INVESTIGATIONS OF THE THERMALLY INDUCED HELIX-COIL TRANSITIONS OF SYNTHETIC DNA

A THESIS

Presented to

The Faculty of the Division of Graduate Studies and Research

By

Adrian L. Oliver

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosphy in the School of Physics

Georgia Institute of Technology

July, 1975

INVESTIGATIONS OF THE THERMALLY INDUCED

HELIX-COIL TRANSITIONS OF SYNTHETIC DNA

Approved:

Roger Wartell, Chairman

Ronald F. Fox

David B. Dusenberry

Date approved by Chairman: 7/22/76

ACKNOWLEDGMENTS

First and foremost I thank my wife Patsy and my daughter Shanda for putting up with me and enduring with me the past five years that I have spent working on this thesis. They have suffered much and now hopefully, will finally get some return on their investment. Secondly, I thank my parents and friends for supporting and encouraging my efforts the entire time.

Another type of thanks is due the Tech physics faculty for their efforts to train yet another physicist. I especially thank Dr. Ronald Fox for his help and encouragement with the theoretical aspects of this work.

Last, but certainly not least, I owe a special debt of gratitude to my thesis advisor, Roger Wartell, without whom I could not have done this piece of work. I greatly appreciated the friendly relationship we had during this work as it made long hours and hard work if not fun, at least bearable.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

HISTORICAL BACKGROUND

Introduction

Deoxyribose nucleic acid, DNA for short, is a large complex biomacromolecule which contains the genetic information that composes the overall construction plan for living organisms. It was not until the famous discovery of Watson and Crick[1] in 1953 of the structure of DNA that the mechanism of information storage and transfer began to be elucidated. Since their discovery, there has been a widespread and systematic search for more information about the structure of DNA and its interactions with other biological molecules. Because of the complexity of the molecule and its various interactions with other molecules, the search for answers to a variety of questions has called for the joint efforts of scientists from different disciplines. What used to be a problem primarily for biologists has become also a problem for chemists and physicists. The combining of efforts of different disciplines has led to many physiochemical methods for examining DNA systems and has been very fruitful in terms of information gathered.

In 1958 it was discovered[2] that when a cell divides

and produces two daughter cells, that each daughter cell received the same amount of DNA from the parent cell and further that the parent cell DNA had undergone strand separation to accomplish the division. DNA strand separation requires some method of unwinding process and also some mechanism for the rejoining of the strands in the daughter cells. The process of strand separation has been the subject of much research and is also the topic of this thesis. The interest in strand separation includes both theoretical calculation and a variety of biochemical-physical experiments. This interest has in some cases necessitated the use of synthetic DNA as it is a somewhat simpler system to study. Much current DNA research involves using synthetic DNAs to study the unwinding mechanism in general and more specifically to make comparisons with various statistical mechanical models which attempt to explain the unwinding process. This is done not only to gain insight into the unwinding mechanism but also to evaluate thermodynamic parameters related to DNA structure. In this report I will extend an existing theoretical model for strand separation and test it by comparing experimental results from several different synthetic DNAs. I will also discuss the evaluation of the associated thermodynamic structure parameters.

DNA Structure

The DNA molecule is a complex of two polymer strands interwoven in a helical fashion and held together by intermolecular forces. Each chain in the complex consists of an alternating sugar-phosphate backbone with any of four types of bases, adenine, thymine, guanine, and cytosine attached to the sugar molecules. The bases of one strand are hydrogen bonded to the bases on the other strand. Figure 1 illustrates the DNA molecule schematically and shows the chemical structure of the sugar-phosphate backbone and the four bases as they are attached to the sugar molecule. The bases connecting the two strands obey a strict complementarity rule. Namely, thymine is always found hydrogen bonded to adenine. Figure 2 illustrates the two base pairs with their corresponding hydrogen bonds. These hydrogen bonded base pairs are essential for the stability of the DNA molecule. The hydrophobic interactions due to the stacked arrangement of the base pairs, however, contribute the most to the stability of the molecule. Even with the stacking interactions and the hydrogen bonding, the two strands are not held together as tightly as the individual sugar-phosphate backbones. The bonds holding the backbone together are covalent and require about 60-kcal/mole to be disassociated[3] while only about 8 kcal/mole of base pairs are required to separate the two strands from each other.[4] Biologically, this

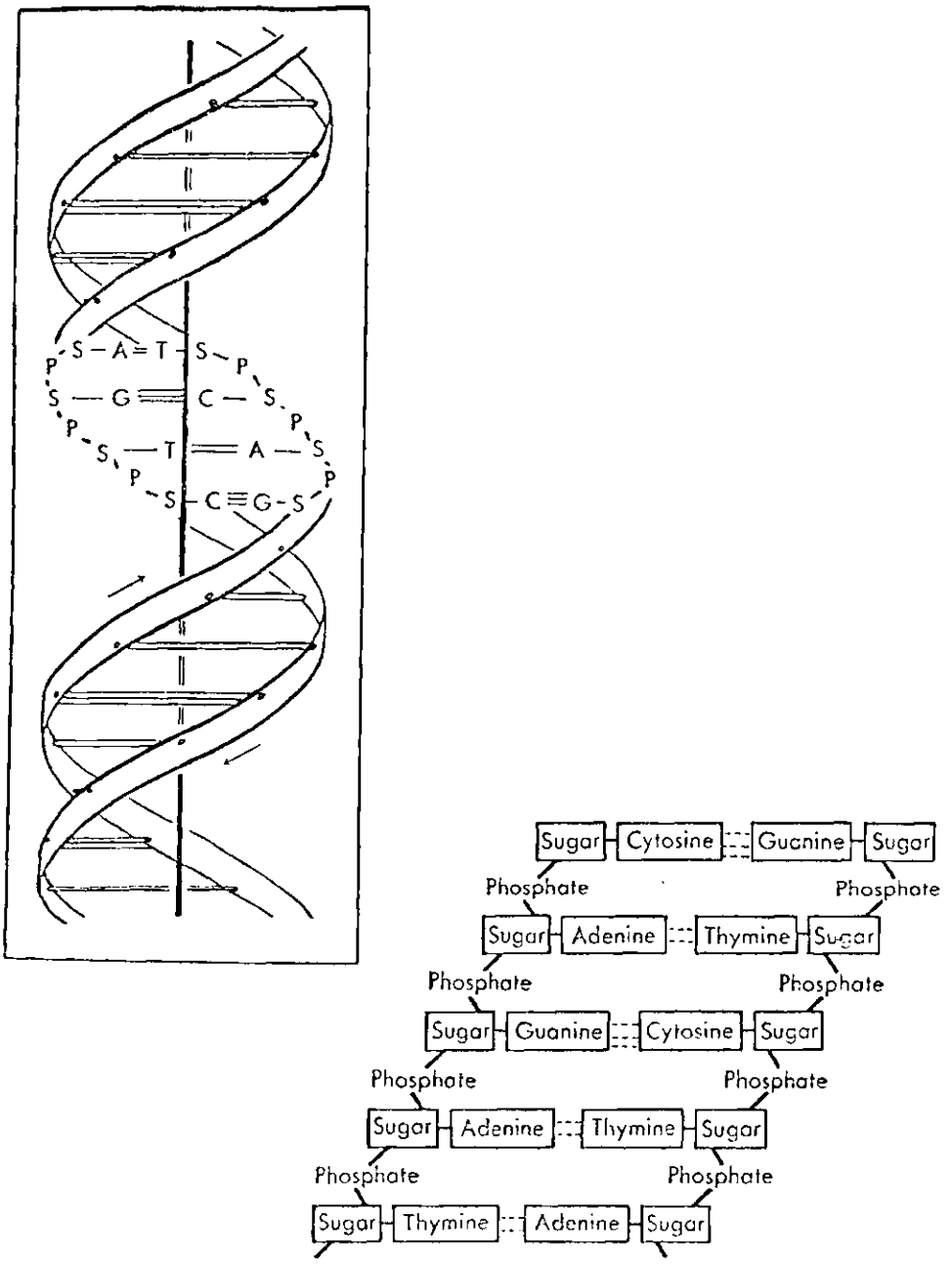
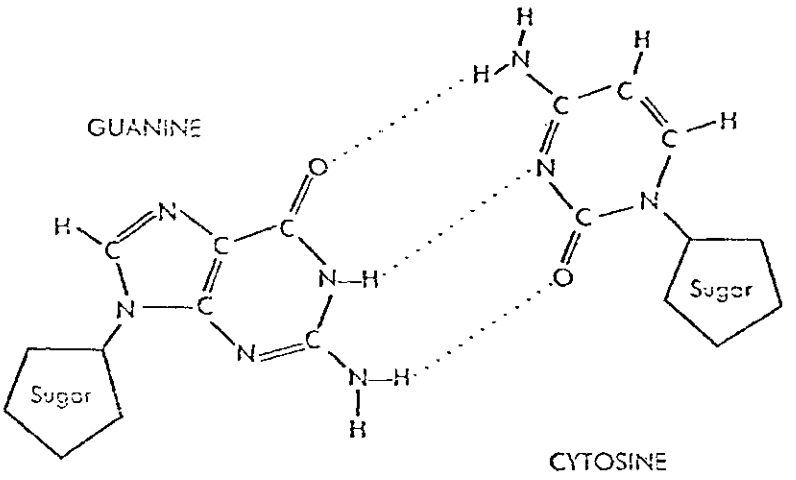
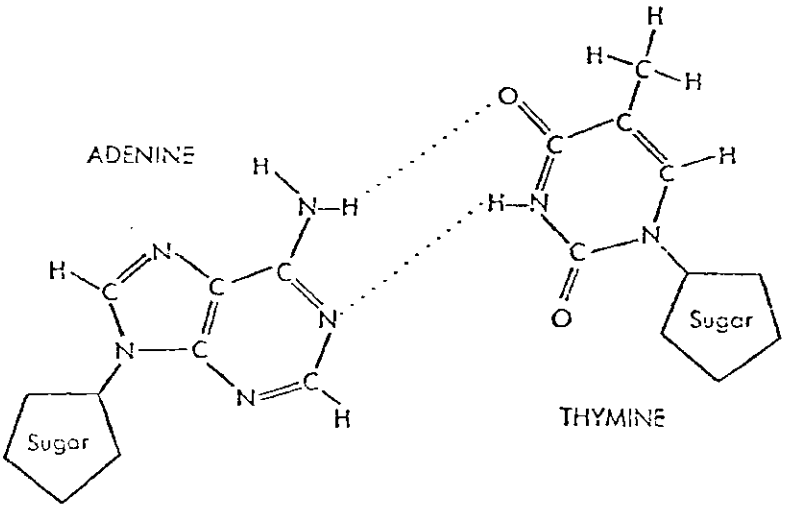


Figure 1. Schematic Representation of DNA.



Structure of guanine and cytosine, showing how they are combined in DNA. The dotted lines are hydrogen bonds.



Structure of adenine and thymine, showing how they are combined in DNA. The dotted lines are hydrogen bonds.

Figure 2. Base Pair Bonding Arrangement.

energy arrangement is very important, DNA duplication is permitted without complete breakdown of the molecule; the strands need only separate to effect a duplication, thus biological integrity of the DNA base pair sequence is maintained due to the complementarity of base pair bonding. Some of the methods for studying strand separation will be discussed in the next and following sections.

Experimental Approaches For Studying Strand Separation

The phenomenon of strand separation has been known by a number of names. Among them are denaturation, unwinding, unzipping, melting, and the helix-coil transition. The helix-coil transition can be brought about in several ways. Heat, high and low pH, low salt concentration, and nonaqueous solvents all have a destabilizing influence. Many physical properties of the DNA are altered in the process, such as optical absorbence and rotation, viscosity, and sedimentation coefficient, making the transition readily accessible to physical measurement. These physiochemical assays have been used to gain information on the molecular bonding and structure of natural and synthetic DNAs.[5-8] Also, helix-coil experiments using synthetic DNAs have been useful in delineating the relative importance of different types of bonding (hydrogen bonding, stacking interactions, etc.) on the stability of DNA in solution.[9-11] One of the most widely used methods is the equilibrium

denaturation by thermal means, with the transition being observed by measuring the ultraviolet absorbence as a function of temperature. Figure 3 illustrates the transition. In the first part of the transition little or no absorbence change is observed as the temperature approaches the melting temperature. This is interpreted as the maintaining of the double-helix structure. As the temperature continues to rise into the melting region, more and more base pairs begin to separate, that is the hydrogen bonds connecting the bases on opposite strands begin to break. Physically, the DNA double-helix is beginning to unwind with loops of unbonded bases being formed in the middle and at the free ends of the molecule. This is observed as a large and sudden increase in the absorbence. The percent increase varies with DNA type but is usually about a thirty percent increase. The sharpness of transition, i.e., the slope in the middle region of Figure 3, is also very dependent on DNA base pair sequence and is approximately ten degrees centigrade wide for natural DNAs and two or three degrees centigrade wide for synthetic DNAs. Natural DNAs exhibit a wider transition than synthetic DNAs because of the random mixing of AT and GC base pairs. The GC base pairs are considerably more thermally stable than AT pairs, thus regions which are lacking in quantities of GC pairs begin to melt out sooner than regions which have a preponderance of GC base pairs. As the temperature continues

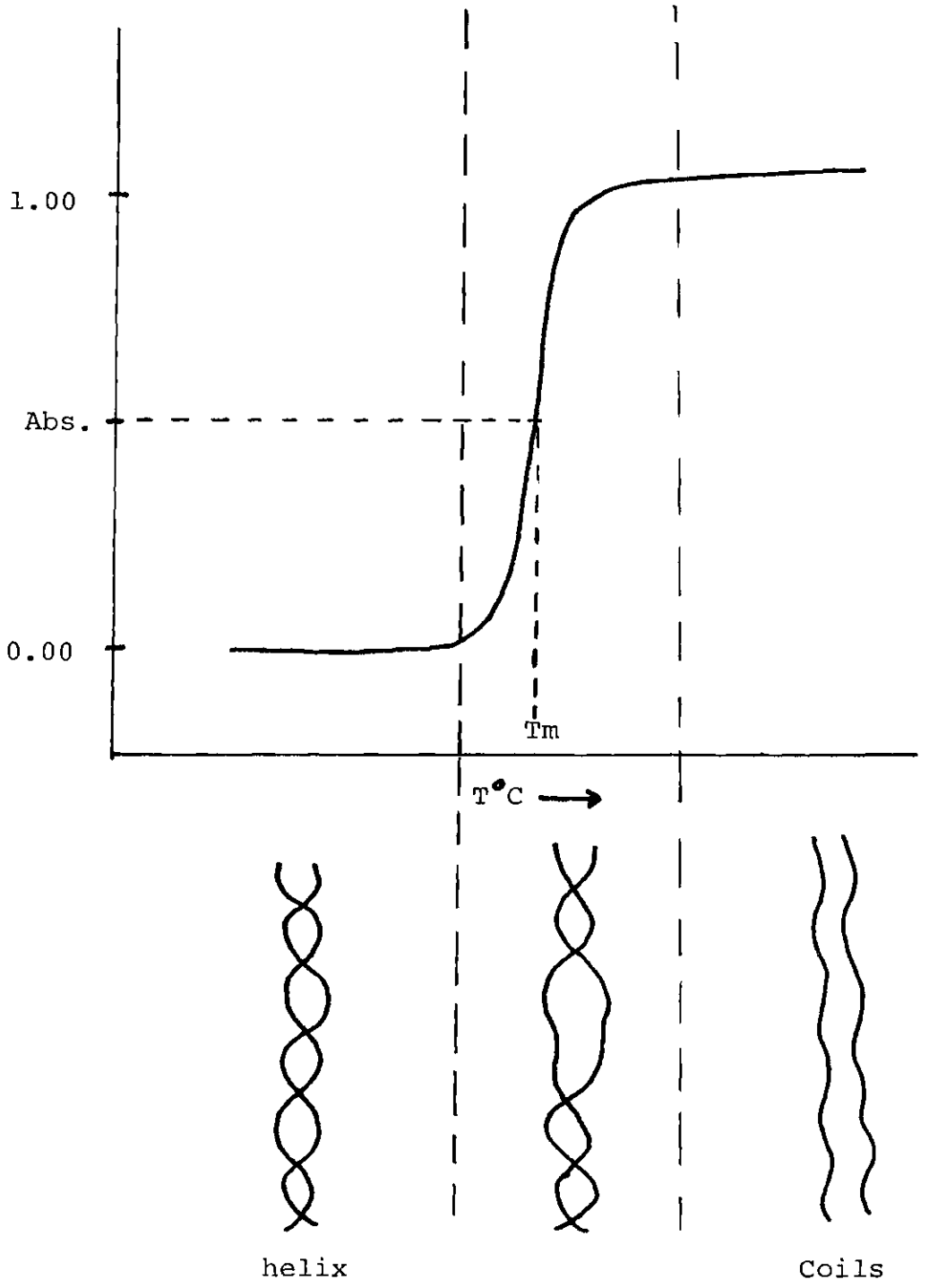


Figure 3. Typical Helix-Coil Transition.

to rise into the third region of Figure 3, the two strands become more and more separated until finally they are completely separated. When this occurs there is no more increase in the absorbence. The transition when observed in this manner is usually completely reversible for synthetic DNAs. Natural DNAs usually do not exhibit the degree of reversibility that synthetic DNAs exhibit. Again, this is due to the random mixing of GC and AT base pairs.

The observation that the ultraviolet absorbence of DNA increases with a temperature increase as shown in Figure 3 is known as hyperchromism. Similar observations have been noted for the temperature dependence of the ultraviolet absorbence of single stranded polynucleotides. The increased uv absorbence reflects a lessened interaction of the bases in the polymer. The optical properties of the polymer depend on the electronic properties of the polymer's sub units and the geometry of the polymer. For polynucleotides this means base sequence and conformation. The absorbence of DNA samples is not a linear function of base composition and hence is dependent on neighbor interactions. The simplest assumption is that the major absorbence hyperchromism results form pair wise interactions of bases on one strand with those on the other strand, not necessarily nearest neighbor interactions. Experimental results and calculations based on simple considerations about the origin of hyperchromism are in substantial agree-

ment. The calculations assume that each stacking interaction between bases or base pairs contributes equally to hyperchromism. For a detailed discussion of this effect see reference 54

Several investigators have used this method to investigate the effect of ionic conditions on the helix-coil transition of synthetic DNAs.[7, 10, 12-14] These studies have indicated that sharp increases in sodium ion concentration can induce major structural changes,[7] and that the width of the melting transition increases with increasing sodium ion content.[12] Wells et al[13] have obtained the following results over a hundred fold salt concentration range: (1) DNAs which have the same base composition but different nucleotide sequences do not show identical helix-coil transitions. (2) For sequence isomeric polymers, the DNA containing both purines and pyrimidines on each of the complementary strands is more stable than the isomer containing only purines on one strand and only pyrimidines on the complementary strand. The single exception to this rule is for d(A) • d(T) and d(AT) • d(AT). (3) All polymers undergo a single helix random coil transition over a narrow temperature range. (4) The transitions are virtually completely reversible, particularly at salt concentrations above .1 M and (5) The melting temperature of all polymers is linear with the logarithm of the sodium ion concentration.

These results were also obtained by Inman and Baldwin[14] with different polymers. Several investigators[15-17] have studied base pair content through thermal denaturations and have all found that the percent GC content varies linearly with the melting temperature and logarithm of the sodium ion concentration. Felsenfeld and Sandeen[18] monitored the thermally induced helix-coil transition at several different wave lengths and used the dispersion of curves obtained to separate from one another the contributions of A-T and G-C base pairs to the denaturation. More recent studies involving the use of the thermally induced helix-coil transitions are directed at comparing the experimental curves of synthetic DNAs to theoretically predicted curves in the hopes of learning more about the actual unwinding mechanism of DNA, evaluating thermodynamic energies involved in stabilizing the helix, and base distribution information. Because these studies are directly tied to theoretical calculations, I will discuss them in the next section with their theoretical counterparts.

Helix-Coil Transition Calcualtions and Experiments

Theory

Most of the theoretical calculations on the equilibrium melting of DNA employ basically the same ideas, whether the DNA in question is a naturally occurring DNA

or a synthetic DNA. Because my work deals specifically with synthetic DNAs only, I will confine this discussion to calculations applying to synthetic DNA. A DNA molecule is considered to be a system of N connected units, each unit being a base pair. Each unit has assigned to it a σ_i value, where the i simply denotes the units position in the molecule. The value of the σ_i's is either +1 or -1, depending on whether the hydrogen bonds connecting the base on one strand to the base on the other strand are intact or broken. The adenine-thymine bond consists of two hydrogen bonds, but in this model they are treated as a unit, where the unit is either intact or broken. A similar situation exists for the three hydrogen bonds connecting guanine and cytosine. This model characterization is equivalent to the well known Ising model of ferromagnetism.[19] In that model, the units consist of electron spins which are pointed up or down and are therefore assigned σ_i values of ± 1 depending on their direction. The partition function for the DNA molecule is calculated by assigning statistical weights to the different possible configurations. These statistical weights are of a parametric nature but can be related to the free energies of base pair bonds and the free energy involved in the stacking conformations of different base pair arrangements. Statistical weights are assigned to groups of intact bonds and to groups of broken bonds. Since groups of broken

bonds can occur in the middle of the molecule as well as at the free ends of the molecule, the assigned statistical weights for the two different kinds of broken bonds can be treated in different ways. Two methods have evolved: (1) Consider interior broken bonds as the same thing as broken bonds at free ends and (2) Consider the two types of broken bonds to be different. The second method is called the loop entropy model and accounts for a difference between the entropy of unbonded strands sandwiched between two helical sections and the entropy of unbonded strands at the end of the molecule. The loop entropy model assigns a statistical weight of $\exp(-L_i)$ to the i^{th} base pair when it is intact and $\exp(+L_i)$ when it is broken. L_i is related to the strength of the hydrogen bonding and also has a associated stacking term. The stacking interaction is expressed as $\exp(U_i, i+1 \sigma_i \sigma_i + 1)$ where $\sigma_i = +1$ if the i^{th} bond is intact and $\sigma_i = -1$ if the i^{th} bond is broken. U is related to the strength of the stacking interaction. An additional entropy factor f(m) is assigned to interior loops with m broken bonds and usually has the form $(\frac{1}{2}m\ 1)^{-k}$ where k is a constant whose value is between 1.5 and 2.0.[20,21,27] The form of f(m) will be discussed in greater detail in Chapter III. The formalism described above is due to Montrol and Goel[22] and has been adopted by others as well.[23-25] Alternative notations have been used by several others with the same ideas being employed

and with similar results.[3,26-29] A discussion of the connection between this model and other statistical models can be found in reference 23.

Experimental

Various researchers have directed their efforts toward comparing experimental melting curves with the theoretical curves which can be generated on the basis of the above calculations with the intent of evaluating the structural parameters used in the theory[10,20,25,30] and to estimate the effect of base pair distribution on the transition curve.[23,31,32] Initial efforts have been very successful, at least to the extent that agreement between theory and experiment has been close if not exact. Wartell[11] obtained qualitative agreement for two synthetic DNAs and was successful in obtaining limits for the parameters used in the theory. One of the problems he and others have had is the difficulty in obtaining highly purified, high molecular weight synthetic DNA. Another problem has been the very limited number of known sequence synthetic DNAs. Recently, newer methods of obtaining long DNA have become available and a wider variety of strand sequences is also available. This has made it possible for me to evaluate structural parameters more precisely and to subject the theoretical model to more severe testing.

DNA Nomenclature And Synthesis

Nomenclature

As an introduction, a brief explanation of polymer nomenclature is in order. Polymer nomenclature is in accord with the IUPAC-IUB recommendations. As an example, consider a double stranded DNA with all adenine bases on one strand and all thymine bases on the other. This is, in long hand, polydeoxyriboadenylic polydeoxyribothymidilic acid. In short hand this becomes poly d(A) • poly d(T), with everything before the dot indicating one strand and everything after the dot representing the other strand. An alternate form is simply d(A) • d(T). Thus d(A-T) • d(A-T) represents a double stranded DNA with alternating adenine and thymine bases on each strand. The polymer nomenclature is designed to show also the anti parallel nature of DNA. Thus d(T-A-C) • d(G-T-A) stands for the polymer which has the deoxy sequence...pTpApCpTpA...in one strand and ...pGpTpApGpT...in the complementary strand, both of the strands written with the 5' - hydroxyl on the left and the 3' - hydroxyl on the right.

Synthesis

There are essentially three methods currently used for making high molecular weight repeating sequence DNAs which are double stranded.

1.) de novo reactions (no template) i.e.: dATP + dTTP $\xrightarrow{\text{DNA Polymerase}}$ d(A-T) • d(A-T) +PPi

2.) Use of short chain deoxyribooligonucleotide templates/primers i.e.: $d(pA)_7 \cdot d(pT)_{11}$

$$dATP + dTTP \xrightarrow{\text{DNA Polymerase}} d(A) \cdot d(T) + PPi$$

3.) Use of preformed polymers as templates/primers i.e.: $d(G\text{-}C) \cdot d(G\text{-}C) + dGTP + dCTP$

$$d(G\text{-}C) \cdot d(G\text{-}C) + PPi \xrightarrow{\text{DNA Polymerase}}$$

The first method has been used successfully to prepare five different DNAs.[33] The composition of the product is a function of several reaction conditions, temperature, change of buffer, pH, ionic strength, and choice of polymerase. Method 2 has also had good success, synthesizing at least thirteen different DNAs.[34] The major disadvantage to this method is that the oligomers must be laboriously synthesized by chemical methods. Method 3 provides a means of amplifying the supply of a polymer once it has been made and characterized. Different DNAs react differently to this method however, which may indicate the relative ability of the DNAs to bind to the polymerase.[35,36] An excellent review of these methods is provided by Wells and Wartell.[37]

After the DNA is synthesized and before it can be used, it must be characterized as to purity, sequence, length, and as to whether equal amounts of the complementary strands are present. Purity can usually be obtained by phenol extractions of proteins[38] and then extensive

dialysis and then checked by uv spectral analysis. Sequence analysis can be obtained by nearest neighbor study[36] and by ultraviolet spectrums, uv absorbence temperature profiles, and circular dichroism profiles.[7,13,39] Length characterization can be accomplished by analytical buoyant density determinations in Cs Cl and in Cs_2SO_4 solutions. Each DNA polymer has a unique and characteristic density value[37] which can serve as a critical means of characterizing subsequent preparations of the same polymer. Sedimentation velocity centrifugation is an extremely sensitive and important method of measuring molecular weight and hence length.[40] Determination of complementary strand concentration can be accomplished by chemical analysis or buoyant density studies.[6] Separation of the strands and then recombination under conditions known to provide a duplex can also be accomplished. Such conditions may be established by means of mixing curves generated by titrating one strand against the other.[37]

CHAPTER II

EXPERIMENTAL EQUIPMENT, MATERIALS AND METHODS

Methods and Equipment

Helix-Coil Transition Curves

Data for the helix-coil curves was obtained by monitoring the ultraviolet absorbence of the DNA solutions as a function of temperature. A Beckman DU monochromator and photomultiplier were used in conjunction with a solid state power supply and digital readout obtained from Update Inc., Madison, Wisconsin. A modified sample compartment and cell holder provided for the uniform heating of the cells. A Lauda circulating temperature bath was employed to heat the cells at a rate of 6 degrees centigrade per hour. Temperature was measured with a platinum resistance thermometer inserted into a solvent cell. A linear bridge provided a millivolt output proportional to temperature. The absorbence and temperature were recorded on a two pen strip recorder.

Melting curves were obtained by averaging three independent helix-coil transitions for each DNA. The DNA concentrations were between .28 and .32 optical density units per cm, measured at the maximum absorbence wave

length. The optical density unit referred to is simply the absorbence measured in a 1 cm path length cuvette. Since absorbence, known as optical density, is equal to the extinction coefficient times the concentration times the path length, the optical density of a sample is a direct measure of the concentration of the sample. The extinction coefficient varies for different DNAs but typically has values around 10^4 $(gram/ml)^{-1}cm^{-1}$. This means that a typical concentration used for melting studies is around 50 micrograms per milliliter. This concentration range is used because it approximates ideal solution concentration and aggregation effects are minimized. Larger concentrations are unnecessary and rarely used. Smaller concentrations may be used when here is a scarcity of material, but greater care and stability of instrumentation is required.

After the DNA was purified by methods to be discussed in section 2.2.2, it was prepared for melting by first recording its absorbence as a function of wavelength at 25 degrees centigrade. This determines the wavelength to be used for melting, i.e., that wavelength where the absorbence is a maximum. It was then centrifuged at 2000 rpm for approximately ten minutes to remove any dust and then carefully piped into a clean, dry, quartz cuvette. It was bubbled with helium gas to remove air bubbles and then the cuvette was stoppered. The helix-coil transition

was monitored continuously from at least fifteen degrees before the transition to at least five degrees after no more absorbence increase was observed. The absorbence readings were tabulated at 0.1 degree centigrade intervals in the transition regions after correction for thermal solvent expansion and plotted according to the expressions:

$$\theta_B(T) = \frac{A(T) - A(\ell)}{A_u - A_\ell} \tag{1}$$

and

$$\frac{d\theta}{dT} = \frac{\theta_B(T_2) - \theta_B(T_1)}{T_2 - T_1} \tag{2}$$

where $\theta_B(T)$ is the fraction of broken base pair bonds at temperature T, A(T) is the absorbence at temperature T, and A_u and A_ℓ are the maximum and minimum sample absorbences respectively. The derivative curve was also used for comparisons with the theory to aid in the evaluation of the theoretical parameters because its shape was much more sensitive to variations in those parameters.

Experimental Equipment

The helix-coil transition was observed by monitoring the ultraviolet absorbence of the DNA as a function of temperature. This experiment calls for the following equipment:

(1) uv spectrophotometer

(2) Container for the DNA sample

(3) Temperature sensing equipment

(4) Temperature control equipment

(5) Recorder

There are many spectrophotometers on the market that come with a wide variety of features. The cost can run from a few thousand to twenty thousand dollars, depending on the built in features one wants. Some of the options include automatic wavelength scanning, automatic cuvette positioner with timing devices for various change rates, built in recorder units, digital display, temperature controls for the cuvette chamber, etc.

I used a Beckman DU spectrophotometer purchased from Update Inc., of Madison, Wisconsin. Update electronically modified the DU to include an external solid state power supply and a digital readout system. Wavelength settings are reproducible to .5 angstroms in the ultraviolet range used. Absorbence measurements with the digital readout are reproducible to $\pm$ .002 over a 60 minute time period. This value is .6 percent of the absorbence range used.

The cuvette chamber was not suited for temperature control and therefore had to be modified. Figure 4 shows the cuvette holder. The chamber was made large enough to accommodate a variety of cuvette holders. The holder is made of brass and has welded to it copper tubing for good

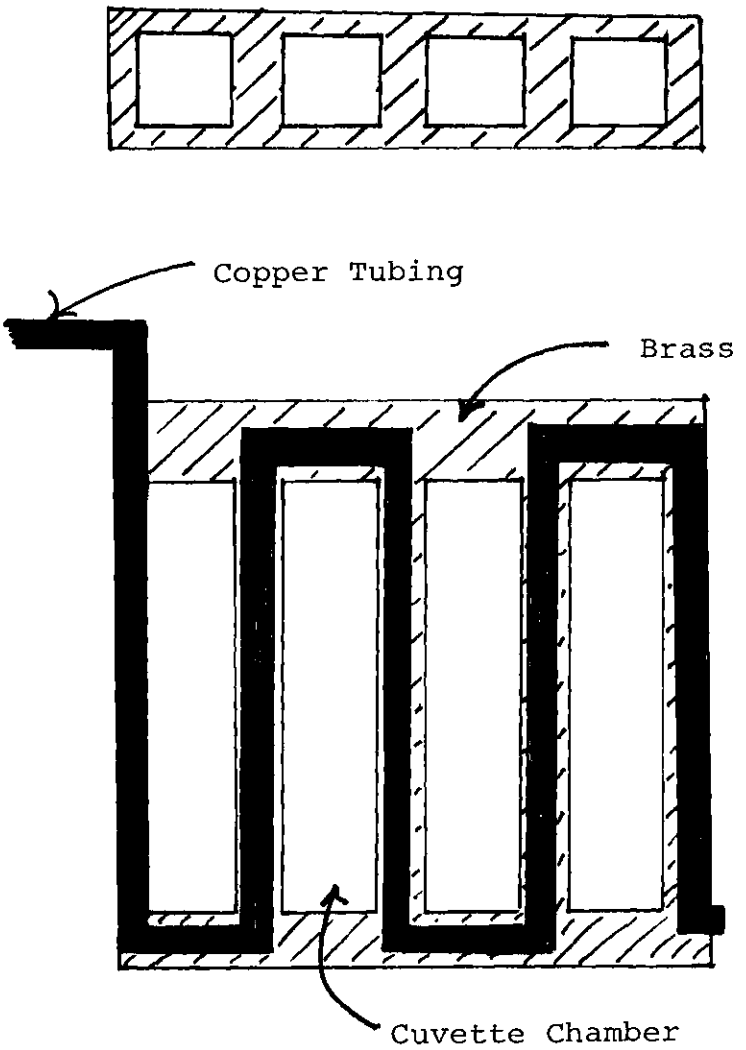


Figure 4. Cuvette Chamber.

thermal contact. Temperature is controlled by passing silicone oil through the copper tubing. The silicone oil temperature is controlled and maintained by a Lauda/Brinkman circulator, which has a control accuracy of ± .01 degrees centigrade. A linear increase of temperature at 6 degrees centigrade per hour was obtained by using a synchronous 1 rpm clock motor attached to the circular temperature control. Temperature of the DNA system was obtained by sealing a Rosemount platinum resistance temperature sensor inside a cuvette which was identical to the cuvette used to contain the DNA. The sensor was connected to a Rosemount linear bridge which had an output of one millivolt per degree centigrade. The variation from a linear output is no more than ± .06 percent in the temperature range from 0.0 degrees centigrade to 100.0 degrees centigrade and the sensor accuracy is to ± .02 degrees centigrade. The linear bridge and the photomultiplier output were wired into an Omniscribe two pen strip recorder to provide for simultaneous recording of the changes in absorbence and temperature. The accuracy of the recorder is rated at ± .3 percent full scale, thus the worst error in temperature was ± .03 degrees centigrade and for absorbence was ± .002. The cuvettes used were 1 ml, 1 cm path length quartz cuvettes purchased from Pyrocell and guaranteed to have the same optical absorbence within 1 percent. In the system

described, there was no discernible difference between cuvettes. A schematic of the entire system is shown in Figure 5.

The entire system was hooked up and checked for proper operation before any data was taken. The DU wavelength adjustment was calibrated using the spectral lines of a mercury lamp. The spectrum of several natural and synthetic DNAs was then taken and compared to their known spectrums with no observable deviations. Proper absorbence readings were obtained by calibrating against absorbence standards purchased from Gilford. The temperature monitoring system was checked against the freezing and boiling points of water and the boiling point of ethyl alcohol at 78.4 degrees centigrade. No deviations were observed.

Materials

Introduction

All of the DNAs used were synthesized by myself with the exceptions of poly d(A-A-T) · poly d(A-T-T) and poly d(G) · poly d(C). d(A-A-T) · d(A-T-T) was synthesized by Ratliff et. al. of Los Alamos Laboratory and generously donated to me. d(G) · d(C) was purchased from P. L. Biochemicals. I chose to synthesize rather than buy because the average length of the commercially available DNA

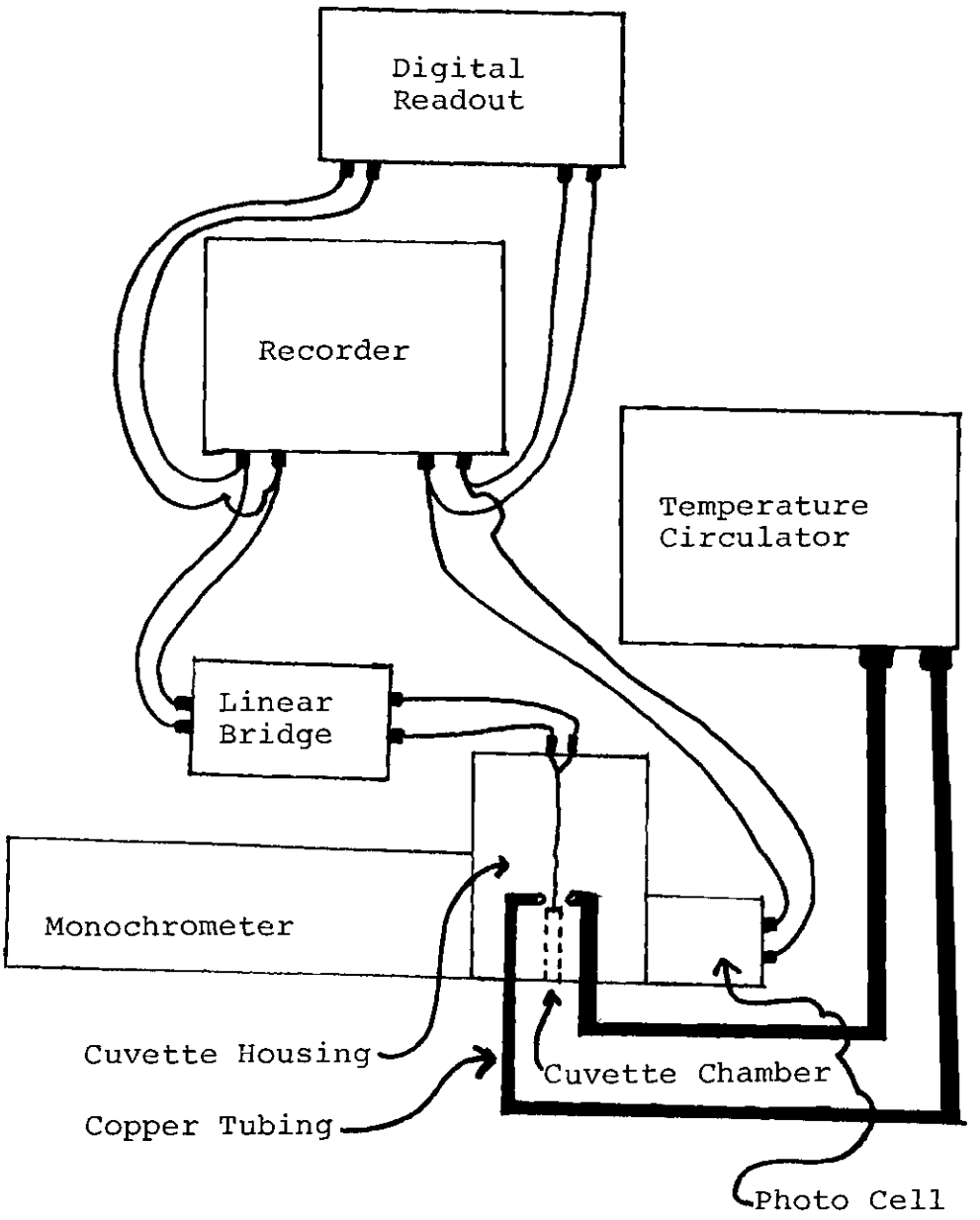


Figure 5. Equipment Schematic.

was approximately one half to two thirds the length I needed. By synthesizing I could obtain a longer DNA by reducing shearing and breakage due to handling and I could obtain sufficient quantities so that the very longest fraction would be enough to do the melting studies. There are several methods available for DNA synthesis. The following paragraph will outline the general procedure and the specifics will come in following sections.

In most cases the reaction ingredients are mixed together in a small test tube and incubated at 37 degrees centigrade for some period of time. The reaction mixtures vary slightly from one preparation to another but always contain different deoxynucleoside triphosphates in amounts whose molar ratios are approximately equal to the molar ratios of the bases in the desired DNA polymer. Various buffers are used to keep the pH in the desired range but usually either sodium or potassium phosphate is used. Virtually all experimenters have found that synthesis proceeds better with small amounts of magnesium chloride and B-mercaptoethanol. Some mixtures include a primer-template consisting of a small amount of the DNA to be synthesized. It has not been found that this is necessary for synthesis, it speeds things up. Lastly, all reaction mixtures contain a DNA polymerase, usually obtained from E. Coli or M. Luteus bacteria. Again, this usually serves to

decrease the reaction time from about twenty hours to about two hours.

AT DNA Synthesis

The deoxynucleoside triphosphates and DNA polymerases used were purchased from P. L. Biochemicals. Poly d(A-T) · poly d(A-T) was synthesized according to the method of Burd and Wells[33] using M. Luteus DNA polymerase. The reaction mixture contained 50 mM potassium phosphate buffered to pH=7.4, 5mM magnesium chloride, 1mM B-mercaptoethanol, 0.8 mM each of the nucleoside triphosphates, 15mM Na_2EDTA, 10 mM d(A-T) · d(A-T) as primer, and 10 units/ml of polymerase. The reaction mixture was kept at a constant temperature of 37 degrees centigrade. The reaction was observed by monitoring the ultraviolet absorbence as a function of time of small aliquots of the main reaction mixture. As the strands form, the uv absorbence decreases, thus the reaction is easy to follow. Initially, the total volume of the mixture was 1 ml, but this was increased to 5 ml and then 10 ml by adding more reaction ingredients to the ongoing reaction. After about a 30 percent decrease in absorbence the reaction was stopped by freezing.

The procedure for synthesizing poly d(A) · poly d(T) is due to Olson, Luk, and Harvey.[42] The reaction mixture contained 67 mM potassium phosphate buffer at pH=7.4, 8.4 mM magnesium chloride, 3 mM B-mercaptoethanol, 4.8 mM

d(A) · d(T) as primer, 1.2 mM each of deoxynucleoside triphosphates, 90 mM reticulomycin, and 5 units of E. Coli DNA polymerase per ml. The reaction is incubated at 37 degrees centigrade and observed with the uv absorbence decrease just like d(A-T) · d(A-T). After about 30 percent decrease in absorbence the reaction is stopped by freezing. It is worth noting that for the d(A) · d(T) all the glassware used must be acid cleaned. This is because alternating AT is very likely to form if there happens to be any primer around, even in minute quantities such as may be stuck to cleaned glassware. To insure the formation of only d(A) · d(T) the glassware needs to be acid cleaned. It is for this reason that the reticulomycin is added. Olson, et. al. reported that the drug inhibited the formation of d(A-T) · d(A-T) while having no effect on d(A) · d(T).

After synthesis, the DNAs must be subjected to various chemical and physical purification steps to remove any extraneous materials. The proteins were removed by phenol[38] or chloroform-isoamyl alcohol[43] extractions. Following this step, the DNAs were exhaustively dialyzed against alternating high (0.5 M sodium ion) and low (.01 M sodium ion) salt solutions. The DNA was dialyzed against four changes with each change being 2 liters of solution per four milliliters of DNA. The DNAs were then lyophilized to a convenient volume and allowed to pass through a Bio-Gel column

as per the procedure of Ratliff.[41] The column used had a total length of 80 cm and a diameter of 2.5 cm. The void volume was 90 ml. The particles used were purchased from Bio-Gel Inc. and had the size designation of A-50 m, 100-200 mesh. The eluent used was .05 M triethylammonium bicarbonate. Fractions were collected in 3 ml aliquots. The longest 5 percent was collected, concentrated, and then dialyzed into one of three solvents; 1 mM sodium citrate +0.5 mM Na_2EDTA, .012 M sodium chloride + 3 mM sodium citrate + .5 mM Na_2EDTA, or .045 M sodium chloride + .2 mM sodium citrate + .5 mM Na_2 EDTA. The DNAs were checked for purity at this point by looking at the uv spectrum and comparing it to spectrums of known DNA samples. Verification of sequence comes from the melting profiles, the Tm, and the width at a specified salt concentration. Figures 6 and 7 show the uv spectra of the alternating AT and d(A) · d(T) used for these melting studies.

The length of the DNAs was measured by sedimentation velocity centrifugation. Measurements were performed with a Beckman Model E ultracentrifuge following the procedures of Eigner and Doty.[40] The ultracentrifuge was generously provided by Dr. John Heirholzer of the Center for Disease Control of Atlanta, Georgia. The solvent employed for these runs was 0.2 M sodium chloride + 0.5 mM Na_2EDTA at a pH=7.2. $S_{20,w}$ values of 16.6, 17.3, and 15.5 svedbergs were ob-

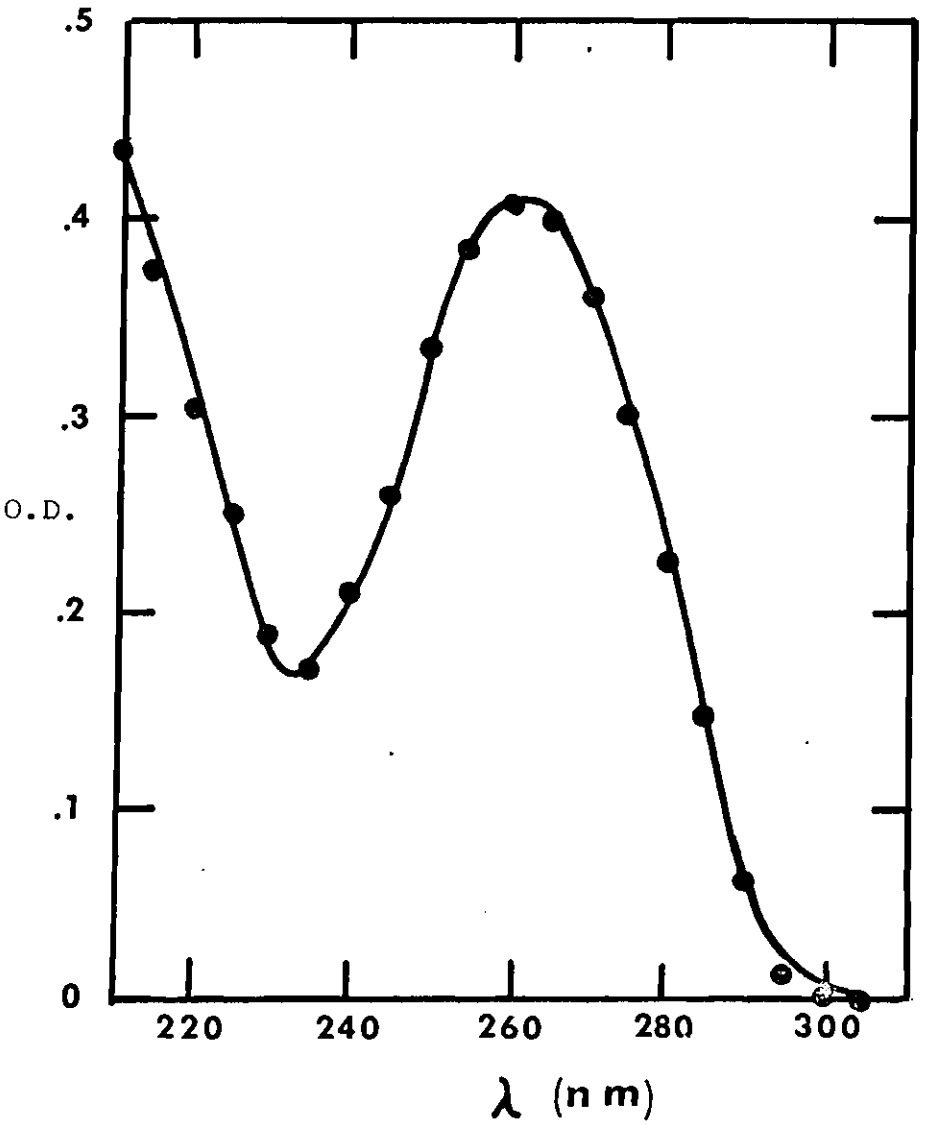


Figure 6. UV Spectra of d(AT) · d(AT).

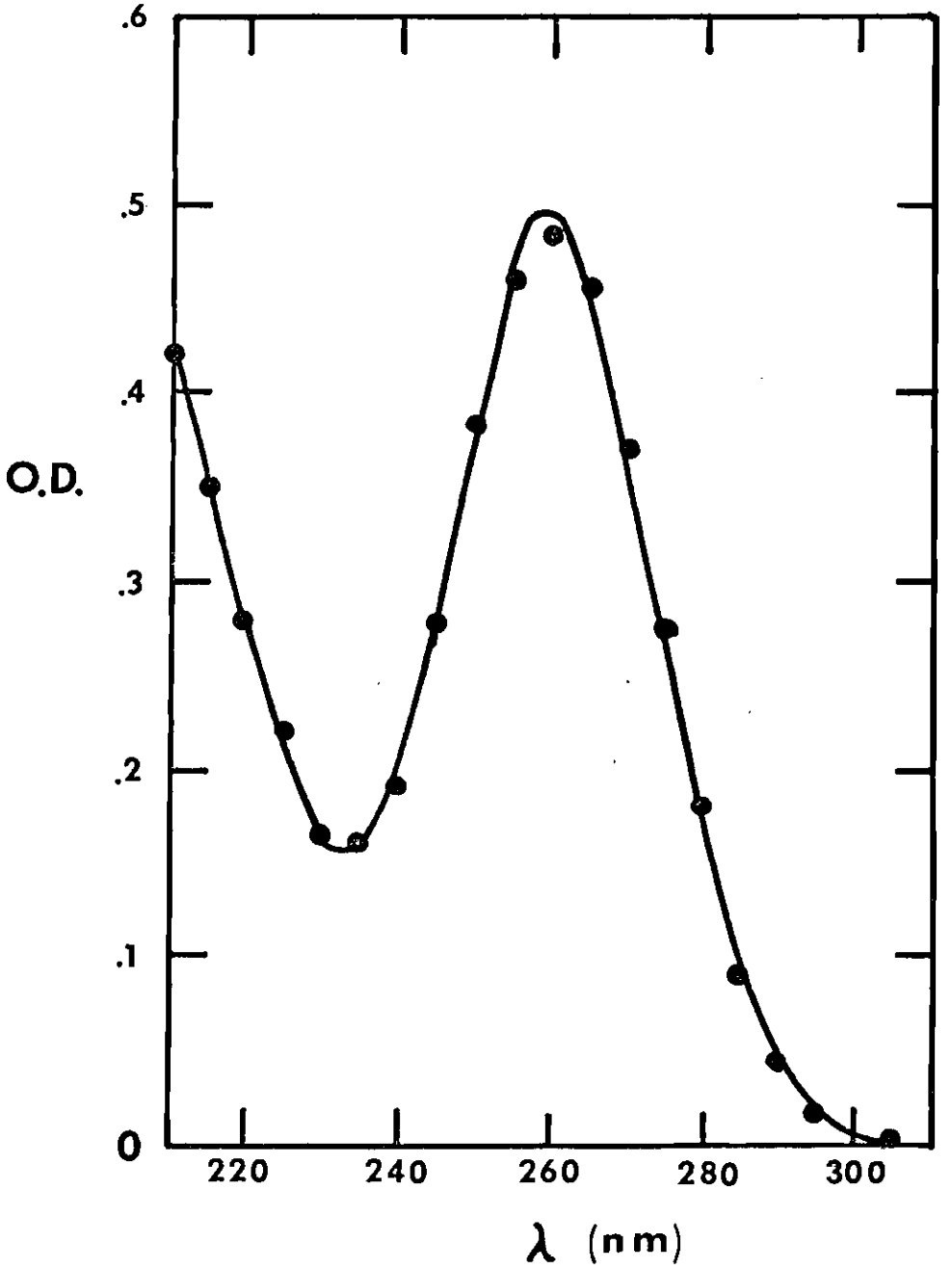


Figure 7. UV Spectra of d(A) · d(T).

served for d(A) · d(T), and d(A-A-T) · d(A-T-T) respectively. These $S_{20,w}$ values indicate duplex lengths exceeding 4000 base pairs in each case.

GC DNAs

There has been no quantitative effort put forth to compare the melting transitions of GC DNAs with theoretical transitions and to evaluate the associated thermodynamic structure parameters. It was felt that an adequate test of the theory would have to be based on an analysis of synthetic DNA polymers containing only G and C bases in a known sequence as well as the AT DNAs. Successful synthesis of poly d(G) · poly d(C)[33] and poly d(G-C) · poly d(G-C)[7,13,33,44] was as early as 1959 for d(G) · d(C) and as late as 1972 for d(G-C) · d(G-C), thus indicating the difficulty involved in their synthesis. I experienced this difficulty for several months while trying all of the methods referenced above with no success whatsoever. The reasons for failure to synthesize a certain DNA when using a proven method are unclear but could easily be attributed to a variety of problems. This is because the precise mechanism of synthesis is not yet known. I finally did synthesize poly d(G-C) · poly d(G-C) but ended up having to buy poly d(G) · poly d(C).

The recipe used for d(G-C) · d(G-C) synthesis was a slight variation of the recipe used by Wells et.al.[13] and

contained 50 mM Tris-H CL at pH=8.3, 1 mM B-mercaptoethanol, 5 mM magnesium chloride, .2 mM each of deoxycytidylic phosphate and deoxyguanosine triphosphate, .25 mM d(I-C) · d(I-C) and 11 units/ml of E. Coli DNA polymerase purchased from P. L. Biochemicals. This was incubated at 37 degrees centigrade for 4 hours with the reaction being followed by watching the uv absorbence decrease of small aligots of the primary batch. After the absorbence was observed to remain constant for a 30 minute interval, the reaction was stopped by freezing. The purification steps used were the same as those used for the AT DNAs. To verify the strand sequence as alternating GC I made uv spectrum and melting temperature comparisons to known alternating GC. The average number of base pairs was 2500 as determined by sedimentation velocity analysis. By placing the DNA on a Bio-Gel column and collecting in fractions as described earlier, I obtained the longest 5 percent which contained on the average over 4000 base pairs per strand as determined by sedimentation velocity analysis. The uv spectrum is shown in Figure 8.

As mentioned above, the poly d(G) · poly d(C) was purchased from P. L. Biochemicals. To obtain both a long duplexed polymer and a one to one correspondence between guanine and cytosine, it was necessary to first unwind the strands in an alkaline salt solution.[55] The complete

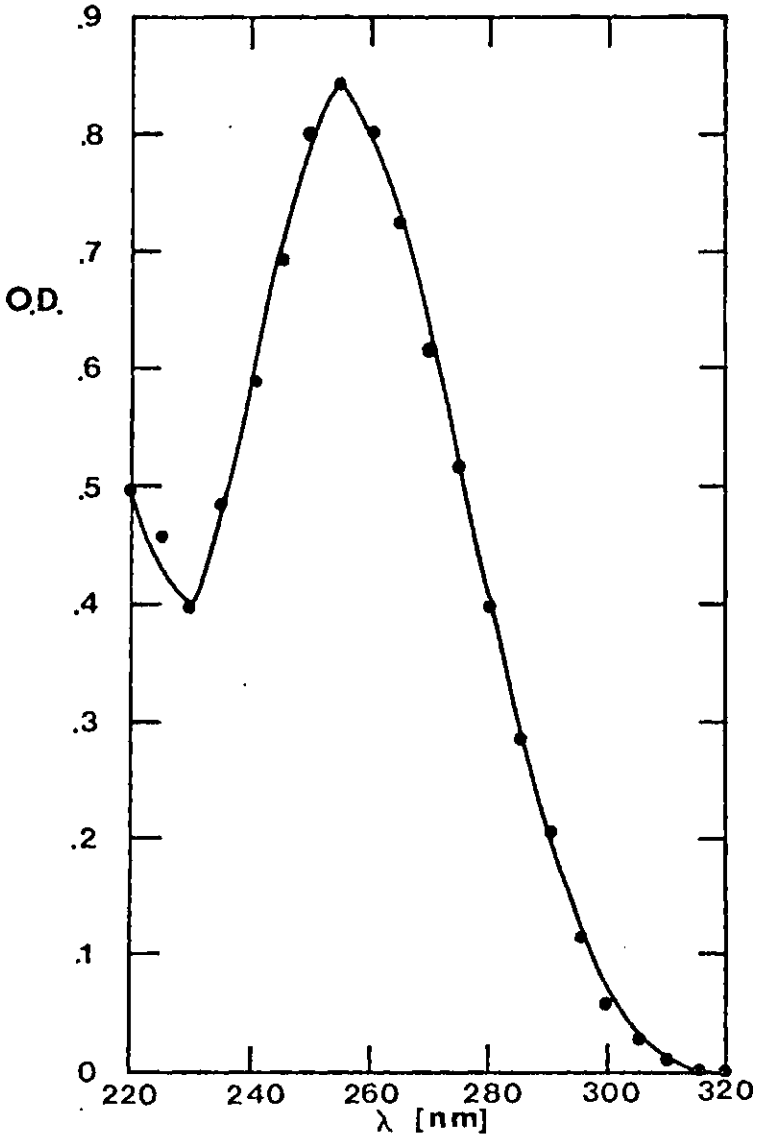


Figure 8. UV Spectra of d(GC) · d(GC).

physical separation of the strands was accomplished by alkaline cesium chloride density gradient centrifugation. In this procedure the density of the cesium chloride is adjusted to be half way between the densities of poly d(G) and poly d(C). High speed centrifugation over a 24 hour period causes the more dense strands to band near the bottom of the centrifuge tube and the less dense strands to band near the top. The separation is usually two to three inches in the centrifuge tube so recovery of one complete band and then the other is accomplished with relative ease. After I had separated the commercially obtained d(G) · d(C) in this manner, I added together the d(G) obtained in this fashion and a quantity of poly d(G) purchased from P. L. Biochemical. The same was done with the d(C). I then placed each of the single strand polymers individually on the Bio-Gel column and collected fractions in units of 3 mls each. Centrifugation velocity analysis of different fractions of each of the polymers provided me with the molecular weights of the different fractions. Duplexing the single stranded polymers in a one to one ratio was accomplished by mixing equimolar amounts of the fractions of each with the highest similar molecular weights in a highly alkaline salt solution. This solution which started at pH=11 was then dialyzed slowly to pH=9, then to pH=8, and finally to neutrality. Centrifugation

velocity analysis yielded an $S_{w,20}$ value equal to 16, which corresponds to over 4000 base pairs. Confirmation of strand sequence and duplexing was accomplished with the uv spectrum and the melting temperature. The uv spectrum is shown in Figure 9. Confirmation of a one to one correspondence between the number of guanine and cytosine bases was not established absolutely, however, results from a circular dichroism analysis and subsequent comparison to Gray's[39] work on base pair sequence determination using circular dichroism techniques indicated that the d(G) · d(C) was good. ("good" in the sense of the above discussion.) This result is significant in that it established the discussed method of obtaining d(G) · d(C) as a relatively quick and easy method of obtaining high molecular weight duplexed d(G) · d(C) with the desired one to one relationship between guanine and cytosine. Figure 10 shows the CD spectrum obtained with the d(G) · d(C) I prepared and the CD spectrum of pure d(G) · d(C) as proposed by Gray.

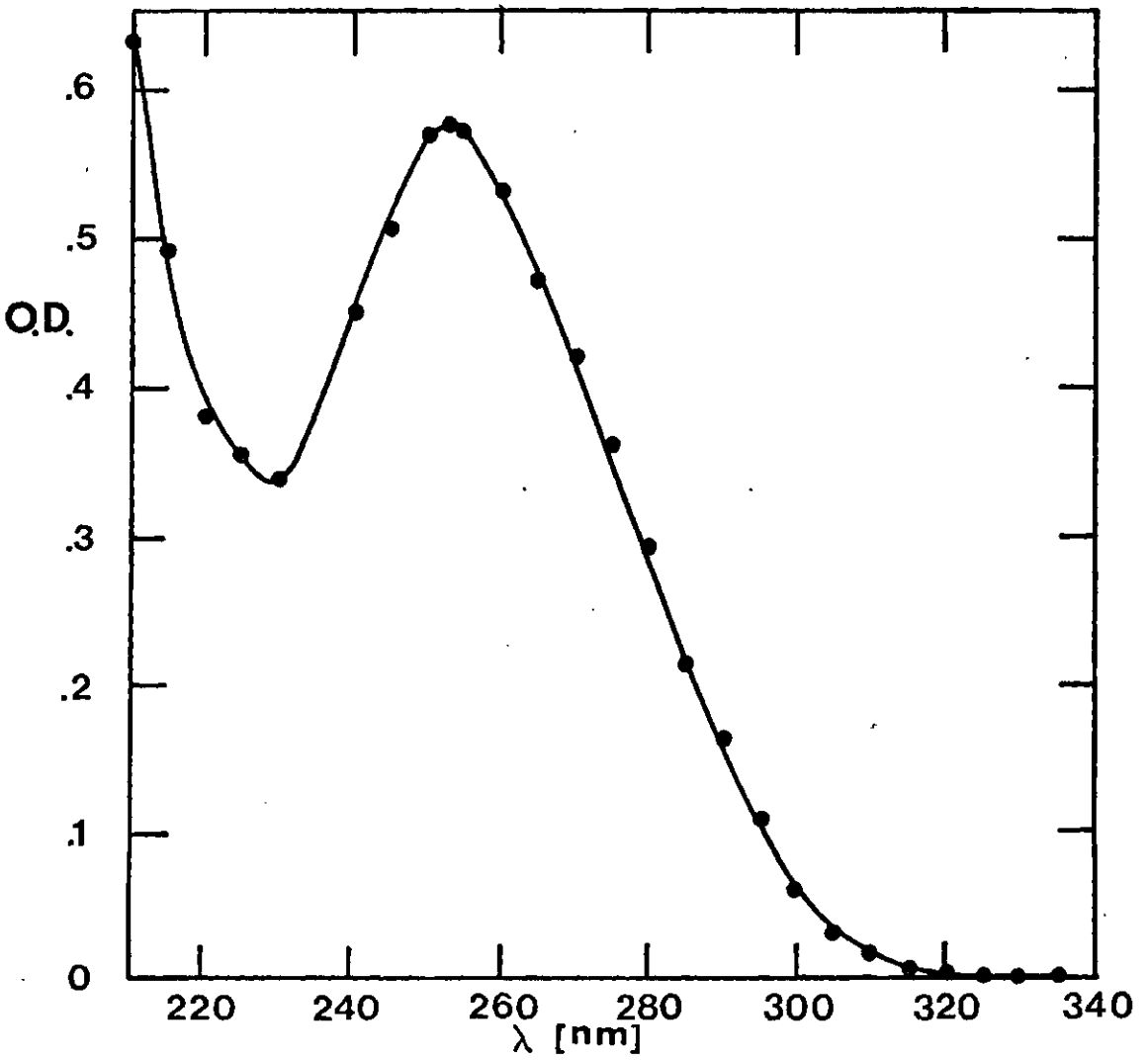


Figure 9. UV Spectra of d(G) · d(C).

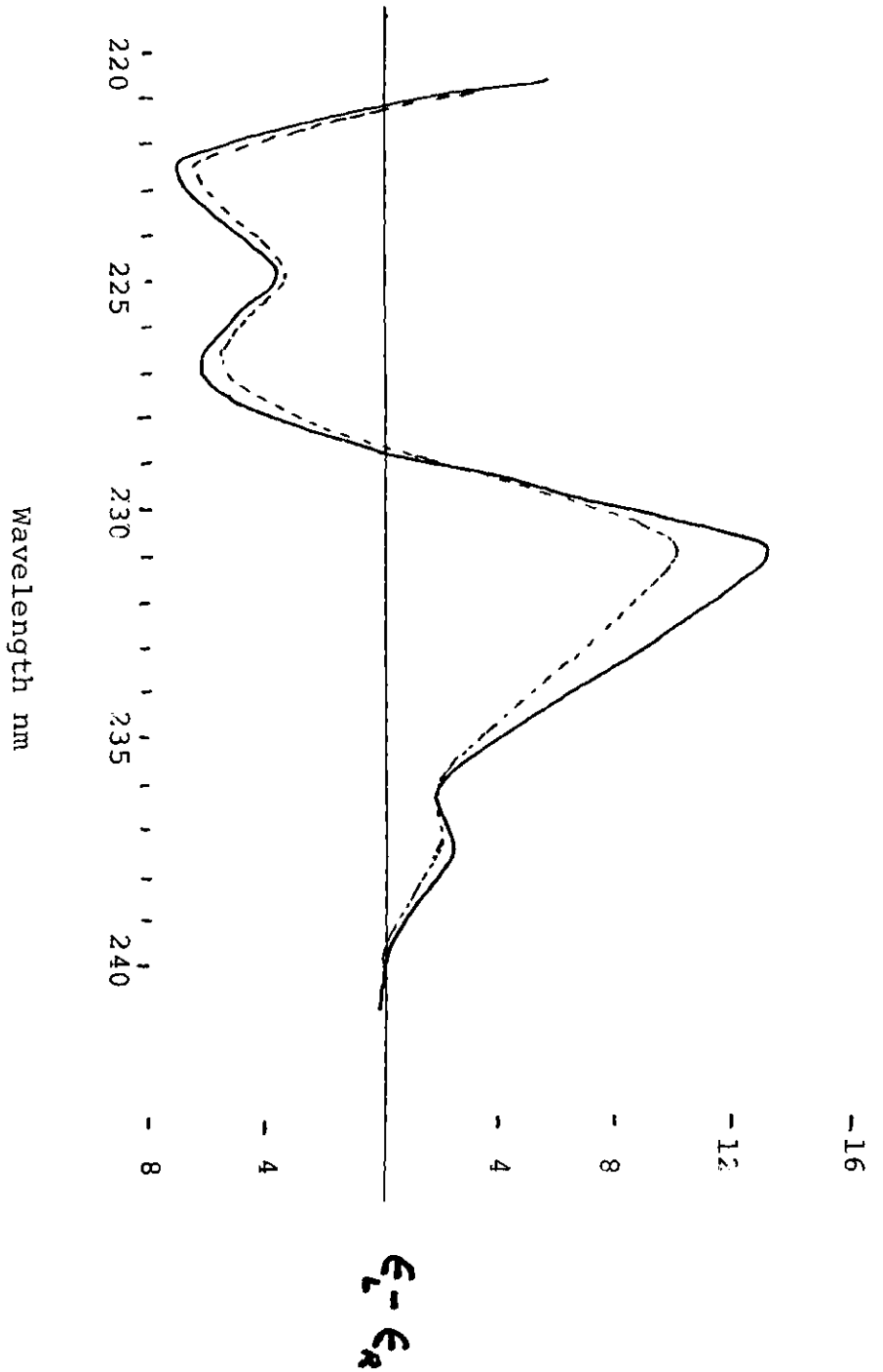


Figure 10. CD Spectra of d(G) · d(C).

CHAPTER III

CALCULATIONS

Basic Theory

The model used to calculate $\theta_B(T)$, the fraction of broken base pair bonds is modified Ising model[19]. We employ the method and notation of Montroll and Goel[19,20]. All of the theory to follow applies equally for AT and for GC DNAs. For continuity and clarity, all of the discussion in this chapter will refer only to DNAs with AT base pairs, with the understanding that the GC theory is obtained merely by replacing A with G and T with C.

Each base pair of a DNA is assumed to be either hydrogen bonded (intact), or non hydrogen bonded (broken). One can represent a base pair mathematically with +1 if the bond is intact and -1 if the bond is broken. The entire DNA molecule can now be represented as a series of bonds, each designated with a σ which can take on a value of $\pm$ 1 depending on the state of the bond. One can specify the configuration of a DNA by the sequence

$$(\sigma_1\ \sigma_2\ \sigma_3\ \text{-----}\ \sigma_N) \tag{3}$$

where

$$\sigma_i = \begin{cases} +1 \text{ if the } i^{th} \text{ base pair is intact} \\ -1 \text{ if the } i^{th} \text{ base pair is broken} \end{cases}$$

Using this notation, the fraction of broken base pair bonds in a molecule with N base pairs is given by

$$\theta_B(T) = \frac{1}{N} \sum_{i=1}^{N} \langle \tfrac{1}{2}(1 - \sigma_i) \rangle \tag{4}$$

where $\langle\ \rangle$ represents the average over all value of the σ_i's. To carry out the averaging process, we need the probability of finding the DNA molecule in a particular configuration, $P(\sigma_1\ \sigma_2\ \sigma_3\ \ldots\ \sigma_N) \equiv P(\sigma)$. $\theta_B(T)$ is then given by

$$\theta_B(T) = N^{-1} \sum_{\sigma} \sum_{i=1}^{N} P(\sigma) \tfrac{1}{2}(1 - \sigma_i) \tag{5}$$

with

$$\sum_{\sigma} = \sum_{\sigma_1=\pm 1} \sum_{\sigma_2=\pm 1} \ldots \sum_{\sigma_N=\pm 1} \tag{6}$$

Using standard procedures of statistical mechanics, the equilibrium probability for finding the DNA molecule in a particular configuration is found by dividing that configuration's statistical weight by the partition function, the

sum over the statistical weights for all possible configurations.

In order to calculate $P(\sigma)$ one must specify the interactions between the base pairs. The modified Ising model makes the following assumptions on the interactions governing DNA stability. The free energy change in breaking the i^{th} base pair depends on the type of base pair (A•T or G•C) and on the type and condition of nearest neighbor (σ_{i+1}) base pairs. A longer range effect is included in $P(\sigma)$ by assigning a 'loop entropy factor' $e(m)$ to loops of m unbonded bases bounded by helical segments. This factor accounts for the difference in free energy between an internal loop of m unbonded bases and an equal number of bases at an open end of the DNA molecule. This difference is mainly due to the configurational entropy difference for the two regions.

Introducing the nearest neighbor interaction and the loop entropy factor, $P(\sigma)$ can be written as

$$P(\sigma) = Z^{-1} \prod_{i=1}^{N} f(\sigma_i, \sigma_{1+1}) \prod_{[m_j]} e(m_j) \qquad (7)$$

where Z is the partition function and is given by

$$Z = \sum_{\sigma} \prod_{i=1}^{N} f_i(\sigma_i, \sigma_{i+1}) \prod_{[m_j]} e(m_j) \qquad (8)$$

where

$$f_i(\sigma_i, \sigma_{i+1}) = \exp[U_{i,\,i+1}\sigma_i\sigma_{i+1} - \tfrac{1}{2}L_{i,\,i+1}(\sigma_i+\sigma_{i+1})] \qquad (9)$$

and

$$e(m_j) = (1 + \frac{m}{2})^{-k} \qquad (10)$$

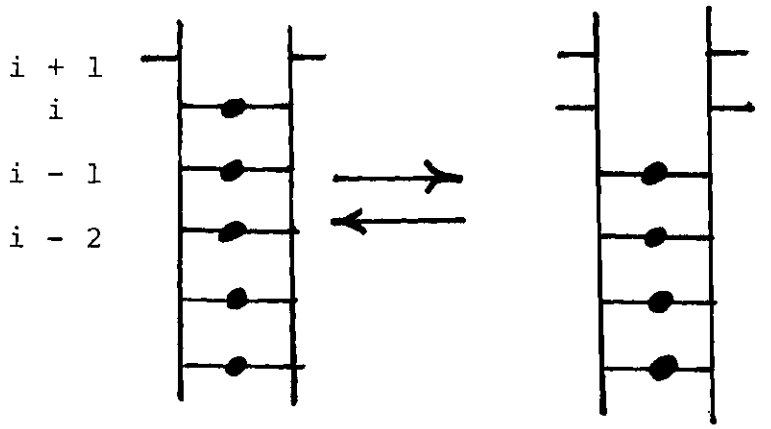
following the example of previous work.[11]

$f_i(\sigma_i, \sigma_{i+1})$ describes the interactions between the i^{th} and $(i+1)^{st}$ base pairs. Equation (9) is the most general form for $f_i(\sigma_i, \sigma_{i+1})$ which includes nearest neighbor interactions since each $\sigma_i^2 = 1$. There is no term included like σ_{i-1} because in summing over the σ_i the nearest neighbor interactions on both sides of a given bond are counted, except at the ends of the molecule. This is accounted for by assuming in the calculation that $f_N(\sigma_N, \sigma_1)$ is equal to one or by allowing N to go to infinity. There is no difference between the results of these two cases.

The $\prod_{[m_j]} e(m_j)$ term in equations (7) and (8) represents the probability of loops of broken base pair bonds that exist in a DNA configuration. $[m_j]$ is the set of loops consistent with a given configuration $(\sigma_1\ \sigma_2\ \ldots\ \sigma_N)$. $e(m_j)$ is related to the probability that the free ends of two joined polymer chains are at the correct distance and orientation for base

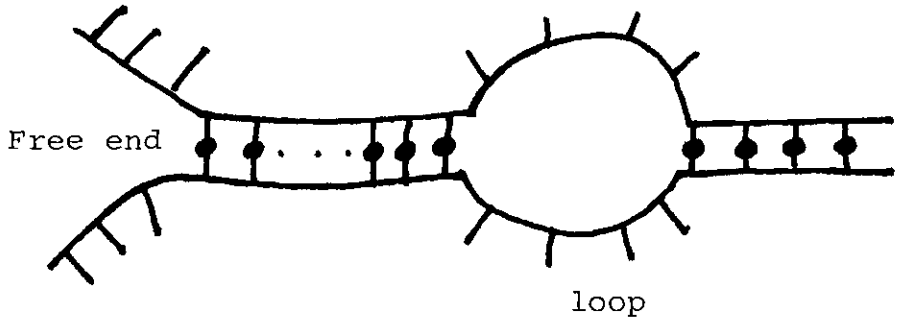
pairing or 'ring closure'. Calculations[9] have shown that for large m, e(m) is proportional to m^{-k}. k equals 1.5 for chains whose links undergo unrestricted random walks[27]. Excluded volume effects are expected to increase k for real polymers. Equation (10) assumes that the configuration entropy of an open end with m unbonded bases is 2^{-k}. Equation (10) also is inaccurate for small loops. This will be of negligible significance for long DNAs. See figure 11b for a schematic of the loop entropy term. $U_{i,\ i+1}$ and $L_{i,\ i+1}$ are parameters related to the free energies of forming base pairs[1]. $U_{i,\ i+1}$ represents the stacking interaction between base pairs. It is equal to $-(4RT)^{-1}$ times the stacking free energy, ΔG_s between the i and (i + 1) bonds. Figure 11a illustrates this important concept. $L_{i,\ i+1}$ is the bond energy parameter. The sum $L_i = (L_{i,\ i+1} + L_{i,\ i-1})/2$ equals $-(2RT)^{-1}$ times the change in free energy, ΔG_{bs}, of forming the i^{th} base pair next to one intact base pair. The stacking interactions between the i^{th} base pair and the i±1 pairs both contribute to G_{bs}. If we label the base pairs $\overset{A}{\cdot}\overset{T}{\cdot}$ and $\overset{T}{\cdot}\overset{A}{\cdot}$ by the numbers 1 and 2 respectively, then $L_{i,\ L+1} = L_{12}$ if the (i, i+1) pair is $\overset{A}{\cdot}P\overset{T}{\cdot}$, $L_{i,\ i+1} = L_{11}$ if the (i, i+1) pair is $\overset{A}{\cdot}P\overset{A}{\cdot}$, etc. The L_{jk} and U_{jk} parameters are related to the more frequently employed Zimm-Bragg[26] parameters by

$$L_{jk} = -\tfrac{1}{2} \log S$$



$$G(i) = G_b(i) + G_s(i,i+1)$$

Figure 11a. Free Energy Change for Broken Bonds.



S_{loop} $S_{free\ end}$ for same number of unpaired bonds

$\exp(- G_{loop}( )/RT$ $-k$

Figure 11b. Free Energy Change for Loop Formation.

and

$$U_{jk} = -\tfrac{1}{4} \log \sigma$$

The U_{jk}'s are assumed to be constant through the transition. They are determined by fitting the theory to the experimental curves. We can evaluate L_{jk} for a one component DNA polymer such as $d(A)_n \cdot d(T)_n$ by expressing it in terms of the enthalpy, ΔH, and entropy, ΔS, changes of stacking a base pair next to an intact pair. For $d(A) \cdot d(T)$ $L = L_{11}$ and from above

$$L = -(2\ RT)^{-1}\ \Delta G_{bs} = -(2\ RT)^{-1}\ (\Delta H - T\Delta S) \qquad (11)$$

R is the gas constant, and T the absolute temperature. At temperatures near the mid point of the helix-coil transition, ΔG_{bs} will be zero and hence

$$\Delta H = T_{cm}\ \Delta S \qquad (12)$$

T_{cm} is not exactly equal to the melting temperature because of the effect of the loop free energy on bond breaking. Assuming $\Delta H/\Delta S$ is constant in the region of transition we obtain

$$L = a(T - T_{cm}) \qquad (13)$$

where

$$a = \Delta H/2\ RTT_{cm} \tag{14}$$

Calorimetric measurements of ΔH and the T_m values from the transition determine L.

Substituting equation (7), (9) and (11) into equation (5) leads to the following expression for the fraction of broken bonds,

$$\theta_B(T) = \tfrac{1}{2}\left(1 + N^{-1} \sum_{\{ij\}} \frac{\partial \log Z}{\partial L_{ij}}\right) \tag{15}$$

where the {ij} in the sum is the number of different types of L_{ij} in the DNA molecule. The derivative melting curve is also useful and is obtained by evaluating the following expression

$$\frac{d\ \theta(T)}{d\ T} = \frac{\theta(T_2) - \theta(T_1)}{T_2 - T_1} \tag{16}$$

To evaluate $\theta_B(t)$ and $\frac{d\ \theta_B}{d\ T}$ it is thus necessary to evaluate the partition function from Equation (8). Z is a function of the arrangement of base pairs along the DNA, the values of the appropriate L_{jk}'s U_{jk}'s and k, and the configurations allowed to partially melted DNAs.

All of the A·T base pairs in $d(A)_n \cdot d(T)_n$ are identical. Therefore one bond energy parameter, and one stacking parameter desribe all $d(A)_n \cdot d(T)_n$ base pairs. We denote these as L_{11} and U_{11}. $d(AT)_n \cdot d(AT)_n$ will also be described by one bond energy parameter L_{12} and one stacking parameter U_{12}. We assume that the stacking interaction is independent of the 3' to 5' base orientation, i.e. $U_{12} = U_{21}$, $L_{12} = L_{21}$. The period three DNA, $d(AAT)_n \cdot d(ATT)_n$ has a repeating sequence of three U's and three L's. We find both L_{12} and L_{11} and U_{12} and U_{11} as shown below.

j,k,ℓ,m

ApTpTpA

. . . .

TpApApT

$L_{j,k} = L_{12}$ $\qquad$ $U_{j,k} = U_{12}$

$L_{k,\ell} = L_{11}$ $\qquad$ $U_{k,\ell} = U_{11}$

$L_{\ell,m} = L_{12}$ $\qquad$ $U_{\ell,m} = U_{12}$

Thus three parameters are needed to determine the melting curves of $d(A)_n \cdot d(T)_n$ and $d(AT)_n \cdot d(AT)_n$; L_{11}, U_{11} and k for $d(A)_n \cdot d(AT)_n$. $d(AAT)_n \cdot d(ATT)_n$ is described by five parameters L_{11}, L_{12}, U_{12}, U_{11} and k. Fitting the predicted curves to the transitions of $d(A)_n \cdot d(T)_n$ and $d(AT)_n \cdot d(AT)_n$ determines U_{11} and U_{12} and provides two independent determi-

nations of k. A comparison can also be made between the experimental melting curve of $d(ATT)_n \cdot d(ATT)_n$ and the curve predicted by the parameters evaluated from the simpler DNAs.

The partition function must sum over all partially bonded configurations of a DNA molecule. Due to their repeating base pair sequence, $d(A)_n \cdot d(T)_n$, $d(AT)_n \cdot d(AT)_n$ and $d(AAT)_n \cdot d(ATT)_n$ can form loops with unequal numbers of bases on the two strands. This is the result of strand sliding. A loop of m bases in $d(A)_n \cdot d(T)_n$ can be distributed in m+1 ways among the two strands. The alternating A-T-A-T ... sequence along each strand of $d(AT)_n \cdot d(AT)_n$ also allows each strand to form single stranded or 'hairpin' loops. Strand sliding and hairpin loops are accommodated in the partition function of $d(AT)_n \cdot d(AT)_n$ by describing loops with m unbonded bases and h helical segments emanating from the loop. h helical segments branching from a loop implies h single stranded sections. The number of distinquishable ways of placing m bases on h sections with at least one base/section is $(m-1)!/(h-1)!(m-h)!$ Strand sliding in $d(AAT)_n \cdot d(ATT)_n$ can occur in three base jumps in each strand. Once the first and last base pairs of a loop are specified, each strand of the loop has restricted number of bases. These restrictions are described in the derivation of the $d(ATT)_n \cdot d(ATT)_n$. Although conformational hindrance along the DNA backbone will not allow every sequence permitted loop, the three DNAs are treated equally.

$$\underline{d(A)_n \cdot d(T)_n}$$

Consider a portion of a DNA molecule undergoing unwinding as shown in Figure 11a. The statistical weight of this configuration of ℓ_1 intact bonds followed by m_1 broken bonds ---, ℓ_n intact bonds, m_n broken bonds can be written as

$$g_1(\ell_1)d_1(m_1)g_2(\ell_2)d_2(m_2)----g_n(\ell_n)d_n(m_n) \tag{17}$$

where the $g_i(\ell_i)$ represent statistical weights of stretches of ℓ_i intact base pair bonds and the $d_i(m_i)$ represent the statistical weight of loops with m_i unbonded bases. From equation (9)

$$g(\ell) = \exp[U_{11} - L_{11})\ \ell - 2U_{11}] \tag{18}$$

where the subscripts on U and L denote the one type of base pair in $d(A)_n \cdot d(T)_n$. The equation for the $d_i(m_i)$ is found by combining equations (9) and (14) and including a strand sliding term to count the number of ways of placing the bases on two strands.

$$d(m) = (m + 1)\ (1 + \tfrac{m}{2})^{-k} \exp[-2U_{11} + \tfrac{1}{2}\, m\ (U_{11} + L_{11})] \tag{19}$$

The partition function for the DNA polymer is found by sum-

ming over all terms such as equation (17) over all possible configurations.

$$Z = \sum_{n=1}^{[\frac{N}{2}]} \sum_{\{\ell_j\}} \sum_{\{m_j\}} \prod_{j=1}^{n} g_j(\ell_j) d_j(m_j) \tag{20}$$

where $[\frac{N}{2}]$ is the largest integer value of $\frac{1}{2}$ N and $\{\ell_j\}$ and $\{m_j\}$ are the sets of ℓ_j's and m_j's consistent with

$$\sum_{j=1}^{n} (\ell_j + \frac{1}{2} m_j) = N, \; 0 < n < [\frac{N}{2}] \tag{21}$$

The sums $\sum_{\{\ell_j\}}$ and $\sum_{\{m_j\}}$ in equation (20) can be replaced with an integral by expressing equation (21) as a delta function and multiplying equation (20) by

$$1 = \exp\{\beta[N - \sum_{j=1}^{n} (\ell_j + \frac{1}{2} m_j) \tag{22}$$

where β is an arbitrary constant.

Using the Fourier integral representation of the delta function one obtains

$$Z = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{N(\beta + i\alpha)} \, d\alpha \sum_{n=1}^{[\frac{N}{2}]} \prod_{j=1}^{n} \phi_j \; \psi_j \tag{23}$$

with

$$\phi_j = \sum_{\ell_j=1}^{\infty} g_j(\ell_j) \exp[-(\beta + i\alpha)\ell_j] = \phi \tag{24}$$

$$\psi_j = \sum_{m_j=1}^{\infty} d_j(m_j) \exp[-(\beta + i\alpha)\frac{m_j}{2}] = \psi \tag{25}$$

Selecting β such that $|\phi\psi| < 1$ and letting $n \to \infty$ for large N

$$Z = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{N(\beta + i\alpha)} [(\phi\psi)^{-1} - 1]^{-1} \, d\alpha \tag{26}$$

Substituting the expressions for $g(\ell)$ and $d(m)$ from equations (18) and (19) into equations (24) and (25) respectively, we find

$$\phi = \frac{\exp(-U_{11} - L_{11})}{\lambda - \exp(U_{11} - L_{11})} \tag{27}$$

with $\lambda = \exp(\beta + i\alpha)$ and

$$\psi = 2^k \, e^{-2U_{11}} \, [r^{-2}\{T(r,k-1) \; 1 \; T(r,k)\} - 2^{-k}] \tag{28}$$

where $r \equiv \lambda^{-\frac{1}{2}} \exp[\frac{1}{2}(U_{11} + L_{11})]$ and $T(r,k) \equiv \sum_{m=1}^{\infty} r^m \, m^{-k}$. β is chosen such that $|r| < 1$ and $T(r, k)$ converges. $T(r, k)$ is known as the Truesdell function. To obtain Z, the inte-

gral in equation (26) is evaluated by the method of residues. Changing the contour to a circle in λ space one obtains

$$Z = \frac{1}{2\pi i} \oint d\lambda \ \lambda^{N-1} \ [(\phi\psi)^{-1} - 1]^{-1} \tag{29}$$

The poles of the integrand are the roots of

$$(\phi\psi)^{-1} - 1 = 0 \tag{30}$$

which are obtained by substitution of equations (27) and (28) into (30) to obtain

$$\lambda - \exp(U_{11} - L_{11}) - \exp(-2U_{11})\{2^k r^{-2}[T(r,k-1) - T(r,k)] - 1\}$$

$$x \exp[-(U_{11} + L_{11})] = 0 \tag{31}$$

Letting $N \to \infty$ equation (29) one now obtains

$$Z \sim \lambda_{max}^N \tag{32}$$

where λ_{max} is the largest root of equation (31) and is obtained numerically as a function of L_{11}. The fraction of broken base pair bonds is thus from equation (15)

$$\theta_B(T) = \tfrac{1}{2}(1 + N^{-1} \frac{\partial \log Z}{\partial L_{11}}) = \tfrac{1}{2}(1 + \frac{\partial \log \lambda_{max}}{\partial L_{11}} \tag{33}$$

d(A-T) · d(A-T)

A detailed calculation for the polymer can be found in reference 25 and will be sketched here for continuity and as an introduction to the integral matrix formalism used for d(AAT) · d(ATT). The calculation is similar to the one just presented for d(A) · d(T) except that alternating AT has two arrangements for an adenine-thymine base pair instead of one; and each strand of alternating AT can fold back on itself to form helical regions. These additional possibilities must be included to properly consider the helix-coil transition.

The two arrangements for an adenine-thymine base pair are $\overset{A}{\underset{T}{\cdot}}$ and $\overset{T}{\underset{A}{\cdot}}$ and will be designated as type 1 and type 2 respectively. Since the first and last bond types of a helical region can be of either type, the weight of a helical region of ℓ base pairs starting with type p and ending with type q will be designated as $g_{pg}(\ell)$. The weight of a loop with m unbound bases and h helical segments branching from it is k(m,h). The helical branches result from hairspinning of single strands. A typical configuration of helical and coil regions such as that shown in Figure 12 will have the weight $g_{12}(\ell_1)\ k(m_1h_1)\ g_{22}(\ell_2)\ k(m_2,h_2)\ldots g_{11}(\ell_n)\ k(m_nh_n)$ The weights $g_{pg}(\ell)$ and k(m,h,) are given in equations (34) through (39).

$$g_{11}(\ell) = g_{22}(\ell) = 0 \qquad \ell \text{ even} \tag{34}$$

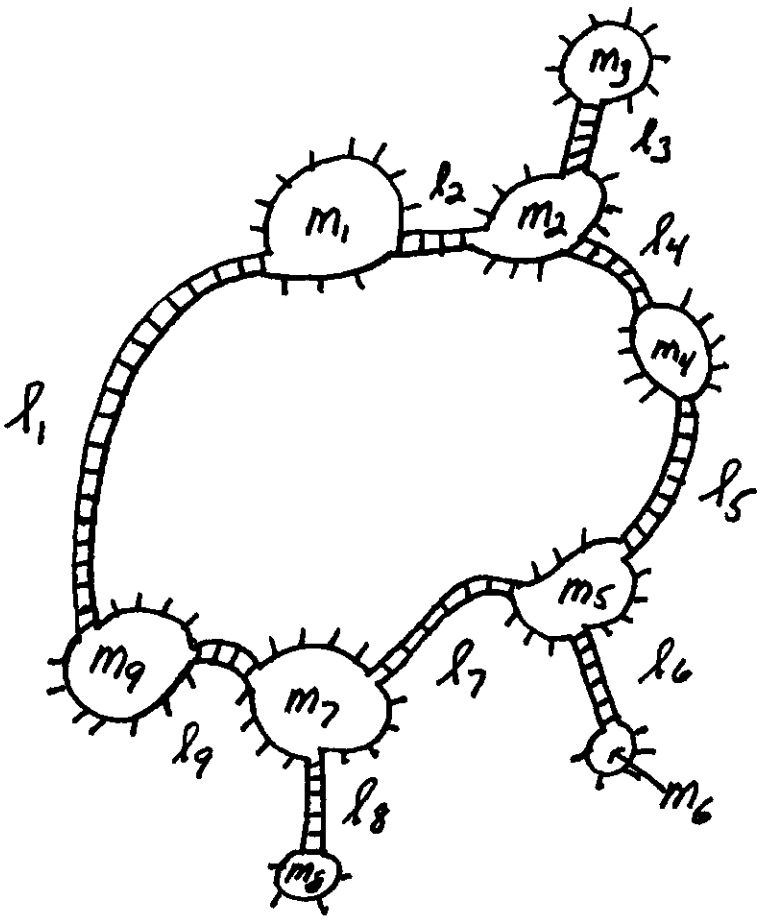


Figure 12. Typical Configuration of Partially Melted DNA.

$$g_{12}(\ell) = g_{21}(\ell) = 0 \qquad \ell \text{ odd} \tag{35}$$

$$K(m,h) = 0 \qquad m \text{ odd} \tag{36}$$

$$g_{11}(\ell) = g_{22}(\ell) = \exp[-2U_{12}+(U_{12}-L_{12})\ell], \quad \ell \text{ odd} \tag{37}$$

$$g_{21}(\ell) = g_{12}(\ell) = \exp[-2U_{12}+(U_{12}-L_{12})\ell], \quad \ell \text{ even} \tag{38}$$

$$K(m,h) = [(m-1)!/(h-1)!(m-h)!](\tfrac{1}{2}m+1)^{-k}$$

$$\times \exp[-hU_{12} + \tfrac{1}{2}(U_{12}+L_{12})m], \qquad m \text{ even} \tag{39}$$

The alternating AT sequence along each strand and the A binding only with T impose the restrictions in equations (34), (35), and (36). Equations (37), (38), and (39) come from equation (9). The factorial part in equation (39) counts the number of distinguishable ways of placing m bases on h single stranded sections of a loop with at least one base on each section. The $(\frac{1}{2}m+1)^{-k}$ is the loop entropy factor while the last part is from nearest neighbor interactions.

To evaluate the partition function, and hence $\theta_B(T)$, the weight factors, equations (34) - (39) must be summed over all configurations possible for the molecule. To write the summation so as to include all possible configurations like equation (17), it is convenient to write the four

possible $g_{pg}(\ell)$'s in matrix form and define G_γ.

$$G_\gamma = \begin{pmatrix} g_{11}(\ell_\gamma) & g_{21}(\ell_\gamma) \\ g_{12}(\ell_\gamma) & g_{22}(\ell_\gamma) \end{pmatrix} \begin{pmatrix} 1 & i \\ 1 & 1 \end{pmatrix} \tag{40}$$

With this notation, the trace of the matrix formed by taking the product of G over all possible ℓ values will give all possible arrangements for the $g_{pg}(\ell)$. Thus the partition function becomes

$$Z = \sum_{n=1}^{[\frac{1}{2}n]} \sum_{\{\ell_\gamma\}} \sum_{\{m_\gamma\}} \sum_{\{h_\gamma\}} \prod_{\gamma=1}^{n} K(m_\gamma, h_\gamma) \; \mathrm{Tr} \left(\prod_{\gamma=1}^{n} G_\gamma \right) \tag{41}$$

where $\{\ell_\gamma\}$, $\{m_\gamma\}$, and $\{h_\gamma\}$ represent the sets of ℓ's, m's, and h's which obey the following two conditions which must hold for all configurations.

$$N = \sum_{\gamma=1}^{n} \left(\ell_\gamma + \frac{m_\gamma}{2}\right) \tag{42}$$

and

$$2n = \sum_{\gamma=1}^{n} h_\gamma \qquad o < n \leq [\tfrac{1}{2}N] \tag{43}$$

Equation (42) counts base pairs and equation (43) follows from the observation that each helical section connects two loop regions. To sum freely over the ℓ's and m's, the integral representation of the delta function,

$$\delta[N-(\sum_{\gamma=1}^{n} \ell_\gamma + \frac{m_\gamma}{2})] \qquad \text{and} \qquad 1 = \exp[\beta(N - \sum_{\gamma=1}^{n} \ell_\gamma + \frac{m_\gamma}{2})]$$

is introduced into equation (41). Thus

$$Z = \tfrac{1}{2}\Pi \int_{-\infty}^{\infty} d\alpha \; e^{N(\beta+i\alpha)} \sum_{n=1}^{[\frac{N}{\alpha}]} \sum_{\ell=1}^{\infty} \sum_{m=1}^{\infty} \sum_{\{h_\gamma\}} \tag{44}$$

$$x \; \Pi \; K(m_\gamma, h_\gamma) \; e^{-\frac{m_\gamma}{2}(\beta+i\alpha)}$$

$$x \; Tr \prod_{\gamma=1}^{n} G_\gamma \; e^{-\ell_\gamma(\beta+i\alpha)}$$

where

$$\sum_{\ell=1}^{\infty} \equiv \sum_{\ell_1=1}^{\infty} \sum_{\ell_2=1}^{\infty} \cdots \sum_{\ell_n=1}^{\infty}$$

To sum freely over the set of h's, equation (43) is introduced as a delta function in the same manner as for the ℓ's and the m's, which after taking the ℓ, m, and h sums inside the product yields

$$Z = \tfrac{1}{2}\Pi \int_{-\infty}^{\infty} d\alpha \; e^{N(\beta+i\alpha)} \sum_{n=1}^{[\frac{N}{\alpha}]} \tfrac{1}{2}\Pi \int_{-\infty}^{\infty} d\,\xi \tag{45}$$

$$x \; e^{-2n(\gamma+i\xi)} \; \psi^m \; Tr(\phi)^n$$

where

$$\psi = \sum_{m=1}^{\infty} \sum_{h=1}^{m} K(m,h) \lambda^{-\frac{m}{2}} \varepsilon^h \tag{46}$$

$$\Phi = \begin{pmatrix} \phi_{11} & \phi_{21} \\ \phi_{12} & \phi_{22} \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \tag{47}$$

$$\phi_{i\alpha} = \sum_{\ell=1}^{\infty} g_{i\alpha}(\ell)\ \lambda^{-\ell} \quad (i,j=1,2) \tag{48}$$

and

$$\lambda = e^{\beta+i\alpha} \quad \varepsilon = e^{\gamma+i\xi} \tag{49}$$

Notice in equation (46) that in summing freely over the h's the sum extends only from 1 to m. This results from assuming at least one unbonded base between helical branches of a loop. The components of Φ are obtained from equations (34), (35), (37) and (38) and are given by the following expressions

$$\phi_{11}=\phi_{22}=[\lambda\exp(-U_{12}-L_{12})]/(\lambda^2-\exp[2(U_{12}-L_{12})]) \tag{50}$$

$$\phi_{12}=\phi_{21}=[\exp(-2L_{12})]/(\lambda^2-\exp[2(U_{12}-L_{12})]) \tag{51}$$

ψ is evaluated from equations (36) and (39) and is given by

$$\psi(E) = [\varepsilon e^{-U_{12}} y^{-1}(1+\varepsilon e^{-U_{12}})^{-3}] \sum_{m=2}^{\infty} y^m (1+\varepsilon e^{-U_{12}})^{2m}/m^k \tag{52}$$

where for convenience y is defined to be $\exp(U_{12}+L_{12})^{\lambda-1}$. One must select β and γ such that $|y|<|$ and $|y(1+\varepsilon e^{-U_{12}})^2|<|$ in order for ψ and ϕ_{ij} to converge. Returning to equation (45)

for Z

$$Z=\tfrac{1}{2}\Pi \int_{-\infty}^{\infty} d\alpha \; e^{N(\beta+i\alpha)} \sum_{n=1}^{\frac{N}{2}} \tfrac{1}{2}\Pi \int_{-\infty}^{\infty} d\xi \; e^{-2n(\gamma+i\xi)} [2(\phi_{11}+\phi_{12})\psi]^n \quad (53)$$

where the identity

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}^n \equiv 2^{n-1} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \quad (54)$$

and the equivalence of $\phi_{11}= \phi_{22}$ and $\phi_{12} = \phi_{21}$ have been used. The rest of the calculation requires no more new physical insight and the interested reader is referred to reference 25. The result is

$$\Theta_B(T)= (1+\lambda_m^{-1}\frac{\alpha\lambda m}{\alpha L_{12}}) \quad (55)$$

where λ_m is the largest root of

$$1 - 8 \lambda \exp(-4U_{12}-L_{12})[2 \sum_{m=1}^{\infty} \frac{y^m}{m^{k-1}} -3 \sum_{m=1}^{\infty} \frac{y^m}{m^k} +y]$$

$$x \; [\lambda-\exp(U_{12}-L_{12})]^{-1} \quad (56)$$

The largest root of equation (56) is found numerically with a computer. The total derivative of equation (56) is taken with respect to L_{12}, the result is solved for $\frac{\alpha\lambda}{\alpha L_{12}}$. The largest root of equation (56) is then placed in that equation,

which generates the theoretical $\Theta_B(T)$ curve as the computer steps through in designated temperature increments.

d(AAT) · d(ATT)

The melting curve of $d(AAT)_n \cdot d(ATT)_n$ will be derived by treating this polymer as a repeating sequence of three different base pair types. The assumption that $L_{12} = L_{21}$ and $U_{12} = U_{21}$ is not made until the effects of the periodicity are accounted for. Consider a configuration of $d(A\text{-}A\text{-}T)_n \cdot d(A\text{-}T\text{-}T)_n$ as an alternating sequence of intact and broken bonds. Call ℓ_j the number of intact base pairs in duplex segment j and m_j the number of unbonded bases in the j^{th} loop of broken bonds. End effects are not important in this calculation since N eventually becomes infinitely large. We will denote $g_{pq}(\ell_1)$ as the statistical weight (s.w.) for ℓ_1 intact pairs in segment 1 starting with an p type pair and ending with a q type pair. $f_{rs}(m_1)$ is the s.w. for a loop of m unbonded bases in loop 1 where the first unbonded base is part of an r type base pair, and the last unbonded base is part of a s type pair. The base pairs for $d(A\text{-}A\text{-}T)_n \cdot d(A\text{-}T\text{-}T)_n$ will be numbered as follows

```
1 2 2'  1 2 2'  1 2 2'
A T T   A T T   A T T
. . .   . . .   . . .
T A A   T A A   T A A
```

The s.w. for one possible DNA configuration is given by

$$g_{11}(\ell_1)\ f_{22'}(m_1)\ g_{11}(\ell_2)\ f_{22'}(m_2)\ \cdots\ g_{11}(\ell_n)\ f_{22'}(m_n) \quad (57)$$

In this configuration the first duplex segment starts with a type 1 base pair and ends with a type 1 base pair. The first loop starts with a type 2 base pair and ends with a type 2' base pair. The ℓ_j and m_j of equation (57) and all other possible configurations must obey the equation

$$\sum_{i=1}^{n} (\ell_j + m_j/2) = N,\ 0 < n < [N/2] \quad (58)$$

[N/2] is the integer closest to N/2. The partition function is given by

$$Z = \sum_{n=1}^{[\frac{N}{2}]} \sum_{[\ell_j]} \sum_{[m_j]} \prod_{j=1}^{n} q_{pq}(\ell_j) f_{rs}(m_j) \quad (59)$$

The sets of $[\ell_j]$ and $[m_j]$ are subject to the restrictions of equation (59). The pg and rs indices are subject to restrictions imposed by the DNA periodicity. We will first treat the restrictions on $[\ell_j]$ and $[m_j]$.

Inserting $\delta[N - \sum_{j=1}^{n} (\ell_j + \frac{1}{2}m_j)]$ and $1 = \exp[N - \sum_{j=1}^{n}$ $(\ell_j + \frac{1}{2}m_j)]$, where β is an arbitrary constant into equation (59) allows the sums over $[\ell_j]$ and $[m_j]$ to extend to infinity:

$$Z = \sum_{n=1}^{[\frac{N}{2}]} \sum_{\{\ell_j\}} \sum_{\{m_j\}} \delta[N - \sum_{j=1}^{\infty} (\ell_j + \tfrac{1}{2}m_j)] \exp \beta[N - \sum_{j=1}^{n} (\ell_j + \tfrac{1}{2}m_j)] \quad (60)$$

$$\prod_{j=1}^{n} g_{pg}(\ell_j)\, f_{rs}(m_j)$$

$$\sum_{\{\ell_j\}} \equiv \sum_{\ell_1=1}^{\infty} \sum_{\ell_2=1}^{\infty} \cdots \sum_{\ell_n=1}^{\infty} \quad \text{and} \quad \sum_{\{m_j\}} \equiv \sum_{m_1=1}^{\infty} \sum_{m_2=1}^{\infty} \cdots \sum_{m_n=1}^{\infty}$$

Writing the delta function as a fourier integral, we can write

$$Z = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\alpha \sum_{n=1}^{\infty} \sum_{\{\ell_j\}} \sum_{\{m_j\}} e^{i\,\alpha[N - \sum_{j=1}^{n} (\ell_j + \frac{1}{2}m_j)]} \quad (61)$$

$$\times\, e^{\beta[N - \sum_{j=1}^{n} (\ell_j + \frac{1}{2}m_j)]} \prod_{j=1}^{n} g_{pg}(\ell_j) f_{rs}(m_j)$$

Rearranging equation (61) to take the ℓ_j and m_j terms inside the product we obtain

$$Z = \tfrac{1}{2}\pi \int_{-\infty}^{\infty} d\alpha \sum_{n=1}^{[\frac{N}{2}]} e^{(\beta + i\alpha)N} \sum_{\{\ell_j\}} \sum_{\{m_j\}} \prod_{j=1}^{n} g_{pq}(\ell_j) e^{-(\beta + i\alpha)\ell_j} \quad (62)$$

$$\times\, f_{rs}(m_j)\; e^{-(\beta + i\alpha)\,\frac{1}{2}m_j}$$

To proceed further we must return to the dual subscript notation for the $g(\ell)$ and $f(m)$ and consider the periodicity restrictions imposed by the repeating sequence of the DNA. The periodicity restricts the terms of Z in two different ways. For $d(A\text{-}A\text{-}T)_n \cdot d(A\text{-}T\text{-}T)_n$ only certain $g_{pg}(\ell)$ are non-zero. The values of ℓ are restricted by which base pairs, p and q,

start and end a helical segment. This will be true in a similar way for r, s, and m of $f_{rs}(m)$. A second restriction is that the type of base which starts a loop will depend on the type of base pair which ends the previous helix segment. A similar restriction occurs for the first base pair of helix segments. It depends on the last base of the previous loop. We will consider the second type of restriction first by noting that for a DNA with a repeating sequence of three there are nine possible ways for a helical section or loop section to begin and end. This makes the use of 3 x 3 matrices amenable to our purpose. We define $G(\ell)$ and $F(m)$

$$G(\ell) = \lambda^{-\ell} \begin{pmatrix} g_{11}(\ell) & g_{12}(\ell) & g_{12'}(\ell) \\ g_{21}(\ell) & g_{22}(\ell) & g_{22'}(\ell) \\ g_{2'1}(\ell) & g_{2'2}(\ell) & g_{2'2'}(\ell) \end{pmatrix} \tag{63}$$

and

$$F(m) = \lambda^{-\frac{m}{2}} \begin{pmatrix} f_{22'}(m) & f_{21}(m) & f_{22}(m) \\ f_{2'2'}(m) & f_{2'1}(m) & f_{2'2}(m) \\ f_{12'}(m) & f_{11}(m) & f_{12}(m) \end{pmatrix} \tag{64}$$

where $\lambda = e^{\beta + i\alpha}$. The arrangement of the $g_{pg}(\ell)$ and $f_{rs}(m)$ within the matrices is chosen to insure that when the matrices are multiplied together that only possible sequences will evolve. We can now rewrite equation (62) as

$$Z = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\alpha \, \lambda^N \sum_{n=1}^{[\frac{N}{2}]} \sum_{[\ell_i]=1}^{\infty} \sum_{[m_i]=1}^{\infty} [(111)[\prod_{j=1}^{n} G(\ell_j)F(m_j)] \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}] \quad (65)$$

To procede further in evaluating equation (32), one takes the sums over the ℓ_j and m_j inside the matrices, and then multiply all n identical matrices together. The results are

$$Z = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\alpha \lambda^N \sum_{n=1}^{[\frac{N}{2}]} (1\ 1\ 1)[\Phi\Psi]^n \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \quad (66)$$

with

$$\Phi = \begin{pmatrix} \Phi_{11} & \Phi_{12} & \Phi_{12'} \\ \Phi_{21} & \Phi_{22} & \Phi_{22'} \\ \Phi_{2'1} & \Phi_{2'2} & \Phi_{2'2'} \end{pmatrix} \qquad \Psi = \begin{pmatrix} \Psi_{22'} & \Psi_{21} & \Psi_{22} \\ \Psi_{2'2'} & \Psi_{2'1} & \Psi_{2'2} \\ \Psi_{12'} & \Psi_{11} & \Psi_{12} \end{pmatrix} \quad (67)$$

and

$$\Phi_{pq} = \sum_{\ell=1}^{\infty} \lambda^{-\ell} g_{pq}(\ell) \qquad \psi_{rs} = \sum_{m=1}^{\infty} \lambda^{-m/2} f_{rs}(m) \quad (68)$$

One now makes use of the relation between any matrix A and its eigenvalues, and expresses $[\Phi\Psi]^n$ in terms of its eigenvalues and eigenvectors.

$$Z = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\alpha^N \sum_{n=1}^{[\frac{N}{2}]} \sum_{i=1}^{3} \sum_{j=1}^{3} \sum_{k=1}^{3} \delta_k^n \, E_k(i) \, T_k(j) \quad (69)$$

δ_k are the eigenvalues, $E_k(i)$ the components of the eigenvector and $T_k(j)$ the components of the eigenvectors of the transpose of [ ]. Expanding the term in brackets one finds

$$Z = \frac{1}{2\pi} \int_{\infty}^{\infty} d\alpha \ \lambda^N \sum_{n=1}^{[\frac{N}{2}]} \sum_{k=1}^{3} \delta_k^n (1 + D_k) \tag{70}$$

where

$$DK = E_k(1)\ T_k(2) + E_k(1)\ T_k(3) + E_k(2)\ T_k(3) + E_k(2)\ T_k(1)$$

$$+ E_k(3)\ T_k(1) + E_k(3)\ T_k(2)$$

Using

$$\sum_{n=1}^{\infty} \delta_k^n = \frac{1}{1-\delta_k} - 1 = \frac{\delta_k}{1-\delta_k} \tag{72}$$

we find

$$Z = \int_{-\infty}^{\infty} d\alpha \ \lambda^N \sum_{k=1}^{3} \frac{\delta_k}{1-\delta_k} (1 + D_k) \tag{73}$$

By changing the integration variable to λ and employing the theorem of the residues one obtains as $N \to \infty$

$$\log Z = \log \ \lambda^N \max \tag{74}$$

where λmax is the largest root of

$$[1 - \delta_1(\lambda)]\ [1 - \delta_2(\lambda)]\ [1 - \delta_3(\lambda)] = 0 \qquad (75)$$

To evaluate the eigenvalues for equation (75) one must return to the matrix $\Phi\Psi$ and express ϕ_{ij} and Ψ_{ij} in closed forms. It is thus necessary to consider the restrictions which the DNA periodicity imposes on $g_{pq}(\ell)$ and $f_{rs}(m)$. Non-zero values will be

$g_{pq}(\ell)$	pq	for ℓ=	$f_{rs}(m)$	rs	for m=
	11	3n+1		11	2+3n
	12	3n+2		12	4+3n
	12'	3n+3		12'	6+3n
	21	3n+3		21	6+3n
	22	3n+1		22	2+3n or 1+3n
	22'	3n+2		22'	4+3n
	2'1	3n+2		2'1	4+3n
	2'2	3n+3		2'2	6+3n
	2'2'	3n+1		2'2'	2+3n or 1+3n

n = 0,1,2.....

The $g_{pq}(\ell)$ and $f_{rs}(m)$ can be evaluated from the above considerations and the expressions

$$f_i(\sigma_i, \sigma_{i+1}) = \exp[U_{i,i+1}\ \sigma_i\ \sigma_{i+1} - \tfrac{1}{2}L_{i,i+1}(\sigma_i + \sigma_{i+1}) \qquad (76)$$

and

$$e(m) = (1 + \frac{m}{2})^{-k} \tag{77}$$

It is at this point that we assume $L_{12'} = L_{12} = L_{2'1} = L_{21}$ and similarly for U_{ij}. We will also write $L_{22'}$ as L_{11} to form with the $d(A)_n \cdot d(T)_n$ notation. Examples of the $g_{pq}(\ell)$ and $f_{rs}(m)$ are

$$g_{11}(\ell) = e^{-U_{12}} e^{\frac{\ell-1}{3}(U-L)} \tag{78}$$

$$g_{22}(\ell) = e^{-U_{11}} e^{\frac{\ell-1}{3}(U-L)} \tag{79}$$

$$g_{2'2'}(\ell) = e^{-U_{12}} e^{\frac{\ell-1}{3}(U-L)} \tag{80}$$

$$f_{11}(m) = (\frac{m+1}{3})(\frac{m}{2} + 1)^{-k} e^{-U_{12}} e^{\frac{m-2}{6}(U+L)} \tag{81}$$

$$f_{22}(m) = (\frac{m+1}{3})(\frac{m}{2} + 1)^{-k} e^{-U_{11}} e^{\frac{m-2}{6}(U+L)} \tag{82}$$

$$f_{33}(m) = (\frac{m+1}{3})(\frac{m}{2} + 1)^{-k} e^{-U_{12}} e^{\frac{m-2}{6}(U+L)} \tag{83}$$

where

$$U = 2 U_{12} + U_{22}$$

and

$$L = 2\,L_{12} + L_{11}$$

The subscripts "12" refer to the interactions found in $d(A\text{-}T)_n \cdot d(A\text{-}T)_n$ while the subscripts "11" refer to $d(A)_n \cdot d(T)_n$. The first term in the $f_{rs}(m)$'s counts the number of ways of forming a loop of size me. We now substitute the expression for $g_{pq}(\ell)$ and $f_{rs}(m)$ into equation (68) to arrive at expressions for ϕ_{pq} and Ψ_{rs}.

$$\phi_{11} = \lambda^{-1}\, e^{-U_{12}(1-e^{U-L_\lambda-3})-1} \tag{84}$$

$$\Psi_{11}=(\tfrac{2}{2})e^{-U_{12}\lambda^{-1}}[r\,\tfrac{\partial}{\partial r}\,T(r,\tfrac{4}{3},k) + T(r,\tfrac{4}{3},k)] \tag{85}$$

where

$$T(r,a,k) = \sum_{n=o}^{\infty} \frac{r^n}{(n+a)^k} \tag{86}$$

and

$$r = \lambda^{-\frac{3}{2}} e^{\frac{U+L}{2}} \tag{87}$$

Similar expressions exist for all the ϕ_{pq} and Ψ_{rs}. The following matrices result

$$\Phi = \frac{\lambda^{-2}}{\lambda(1 - e^{U-L}\lambda^{-3})} \begin{pmatrix} \lambda^2 e^{-U_{12}-L_{12}} & \lambda e^{-U_{11}+U_{12}-L_{12}} & e^{U_{11}-L_{11}-L_{12}} \\ e^{U_{11}-L_{11}-L_{12}} & \lambda^2 e^{-U_{11}} & \lambda e^{U_{11}-U_{12}-L_{11}} \\ \lambda e^{-L_{12}} & e^{2U_{12}-U_{11}-2L_{12}} & \lambda^2 e^{-U_{12}} \end{pmatrix} \tag{88}$$

$$\Psi = \frac{(\frac{2}{3})^k}{\lambda^3} \begin{pmatrix} \lambda e^{U_{11}-U_{21}+L_{11}} Q(r,2,k) & e^{U_{11}+L_{11}-L_{12}} Q(r,\frac{8}{3},k) & \lambda^2 e^{-U_{11}} Q(r,\frac{4}{3},k) \\ \lambda^2 e^{-U_{12}} Q(r,\frac{4}{3},k) & \lambda e^{L_{12}} Q(r,2,k) & e^{2U_{12}-U_{11}+2L} Q(r,\frac{8}{3},k) \\ e^{U_{11}+L_{11}+L_{12}} Q(r,\frac{8}{3},k) & \lambda^2 e^{-U_{12}+L_{12}} Q(r,\frac{4}{3},k) & \lambda e^{U_{12}-U_{11}+L_{12}} Q(r,2,k) \end{pmatrix} \tag{89}$$

with $Q(r,a,k) \equiv r \frac{\partial}{\partial r} T(r,a,k) + T(r,a,k)$.

The sum $T(r,a,k)$ can be approximated for $0.9 < |r| < 1.0$ by

$$T(r,a,k) = \frac{\Gamma(1-k)}{r^a} (-\log r)^{k-1} + r^{-a} \xi(k,a)$$

where $\Gamma(1-k)$ is the Gamma function $\xi(k,a)$ is the Zeta function of Rieman. $Q(r,a,k)$ is obtained directly from equation (89).

To complete the problem, one multiplies the two matrices Φ and Ψ

$$\Phi\Psi = \begin{pmatrix} \Phi\Psi_{11} & \Phi\Psi_{12} & \Phi\Psi_{13} \\ \Phi\Psi_{21} & \Phi\Psi_{22} & \Phi\Psi_{23} \\ \Phi\Psi_{31} & \Phi\Psi_{32} & \Phi\Psi_{33} \end{pmatrix} \tag{90}$$

and evaluates the eigenvalues, each of which is a function of λ. We substitute these expressions in equation (75) and let a computer numerically find the values of λ which satisfies the equation. The fraction of broken base pair bonds is thus from equation (15) and (74)

$$\theta_B(T) = \tfrac{1}{2}[1 + N^{-1} (\frac{\partial}{\partial L_{12}} + \frac{\partial}{\partial L_{11}} + \frac{\partial}{\partial L_{12}}) \log Z] \tag{91}$$

$$= \tfrac{1}{2}[1 + N^{-1} (\frac{\partial}{\partial L_{12}} + \frac{\partial}{\partial L_{22}} + \frac{\partial}{\partial L_{12}}) \log \lambda^N_{max}] \tag{92}$$

$$= \tfrac{1}{2}[1 + (2 \frac{\partial}{\partial L_{12}} + \frac{\partial}{\partial L_{11}}) \log \lambda_{max}]. \tag{93}$$

CHAPTER IV

RESULTS

Introduction

Experimental Melting Curves

Figure 13 shows an absorbence vs. temperature curve for poly d(AAT) · poly d(ATT). This curve is typical of the melting data for all the DNAs examined. Comparison is made to the theoretical curves after correcting for dilution effects and normalizing the absorbence data using Equation (1). A normalized melting curve is shown in Figure 14. One obtains the melting temperature Tm, defined as that temperature where half of the absorbence change has occurred, and the width of transition T(.5), defined as the temperature interval between 25 percent and 75 percent of the absorbence change from the melting curve. Table 1 summarizes the Tm and ΔT(.5) values for the three DNA polymers in the three solvents examined. A plot of the Tm vs. -log[Na+] for each DNA yields a straight line which is consistent with earlier published results.[13]

The width of the transition for d(A) · d(T) increases slightly with increasing sodium ion concentration while it stays approximately the same for d(AT) · d(AT) and decreases slightly for d(AAT) · d(ATT). The transition

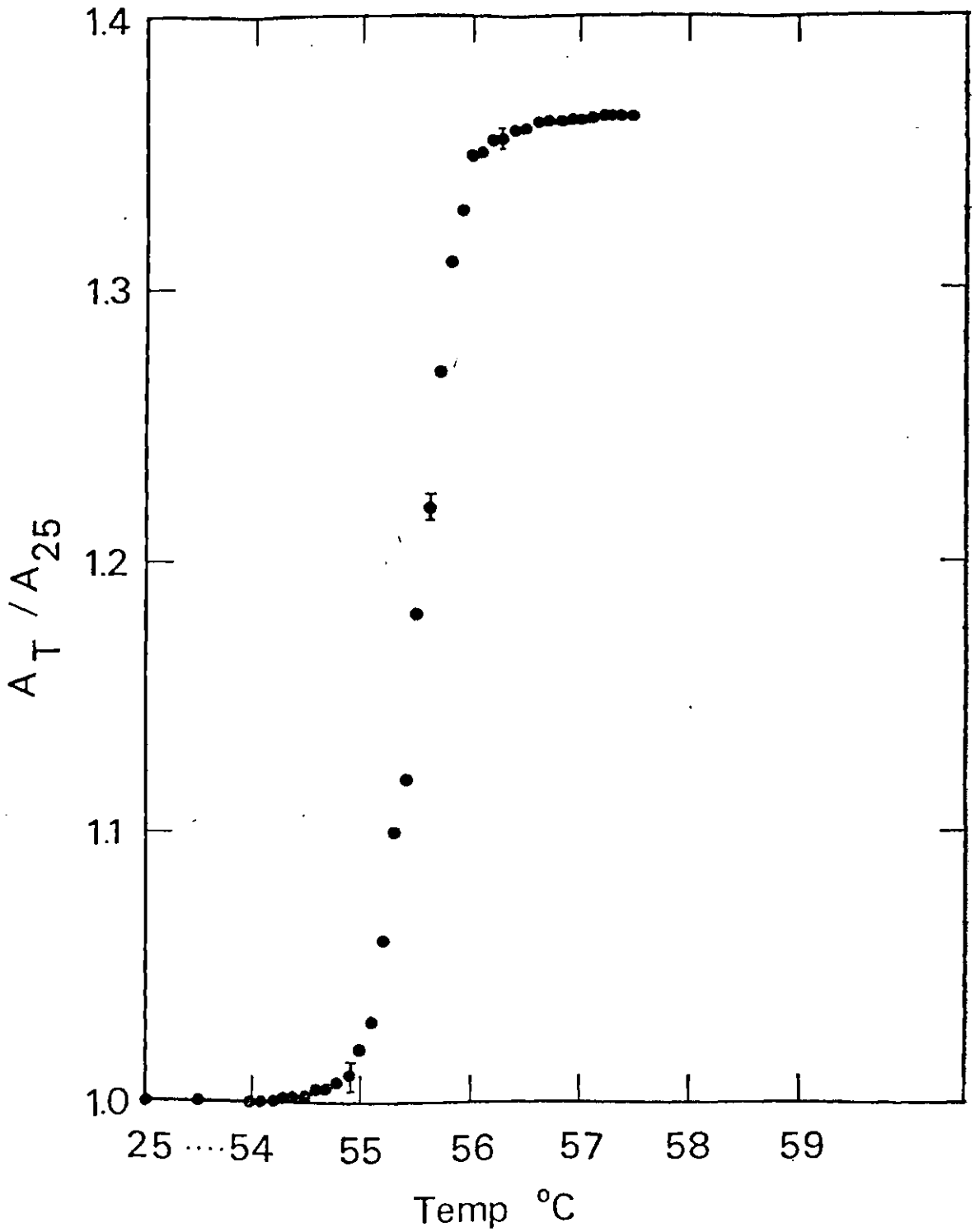


Figure 13. Unnormalized Helix-Coil Data for d(AAT) · d(ATT).

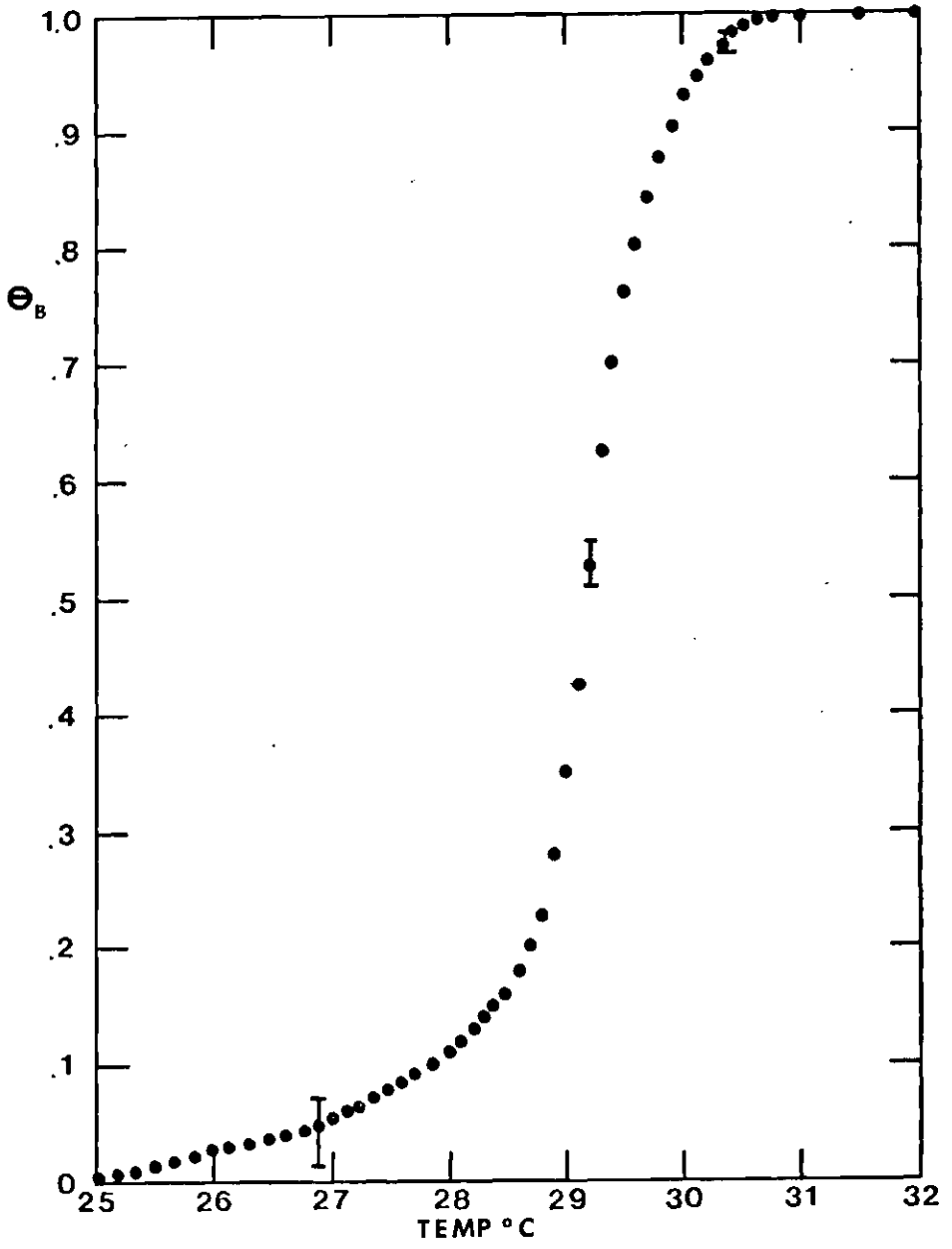


Figure 14. Normalized Melting Profile of d(AT) · d(AT).

Table 1. T_m and $\Delta T(.5)$ Values

	.003M [Na+]		.015M [Na+]		.045M [Na+]	
	T_m	$\Delta T(.5)$	T_m	$\Delta T(.5)$	T_m	$\Delta T(.5)$
d(A) · d(T)	38.10	.40	50.70	.45	59.20	.60
d(AT) · d(AT)	29.20	.60	44.40	.65	54.15	.60
d(AAT) · d(ATT)	32.55	.70	46.40	.60	55.55	.45

width of d(AT) · d(AT) has been shown to increase substantially above 0.1 M salt concentrations.[13] We have not examined this region.

Evaluation of the Thermodynamic Parameters

The theoretical curves were generated by numerically solving Equations (15) and (16) of the preceding sections. The base pair bond parameters, L_{11} and L_{12}, were determined from Equations (10) - (13). The enthalpy change, ΔH, was calculated from Equation (11) using ΔS = -24.8 ev./mole degree K and the Tm value corresponding to the DNA and solvent condition being examined. This entropy value was measured by Scheffler and Sturtevant for d(AT) · d(AT) in 0.01 M sodium concentration.[4] Other experiments indicate this value is almost independent of salt concentration.[9] Theoretical melting curves were generated for d(A) · d(T) and d(AT) · d(AT) by varying k and U_{11} or U_{12} respectively. k was allowed to vary from 1.5 to 2.1 m 0.05 increments. For each k value U_{11} or U_{12} was adjusted to obtain a theoretical curve which most closely matched the experimental melting curve. This procedure sometimes produced more than one pair of values which fit within the experimental error limits. The derivative melting curve was used to select the best U, k values.

AT DNAs

Figure 15 shows the experimental data points for the transition of d(A) · d(T) in .015 M sodium ion. The solid line represents the best fit theoretical curve. It was generated with k=1.55 and U_{11}=2.50. A similar procedure was carried out for d(AT) · d(AT). Shown in Figure 16 and 17 are experimental and theoretical curves in the integral and derivative forms. The parameters which provided the best theoretical curve for d(AT) · d(AT) in 0.015 M sodium ion were k=1.55 and U_{12}=2.40. Parameters evaluated for d(A) · d(T) and d(AT) · d(AT) in 0.003 M and 0.045 M are listed in Table 2. k=1.55 provided the best fit for both DNAs in the three solvents examined. The best U_{11} and U_{12} values varied slightly for the three solvents used. Also listed in Table 2 are the stacking free energies evaluated from the relation between U_{ij} and ΔG_s as given in Equation (9). $\Delta G_s(11)$ is more negative than $\Delta G_s(12)$ in each of the solvents examined. This result is consistent with the observation that d(A) · d(T) has a greater Tm than d(AT) · d(AT).

Theoretical melting curves for d(AAT) · d(ATT) were generated using the parameters evaluated from d(A) · d(T) and d(AT) · d(AT). These curves were compared to the experimental transitions of the period three DNA. This provided a test of the theoretical model. Figure 18 shows the experimental derivative curve and the two theoretical curves

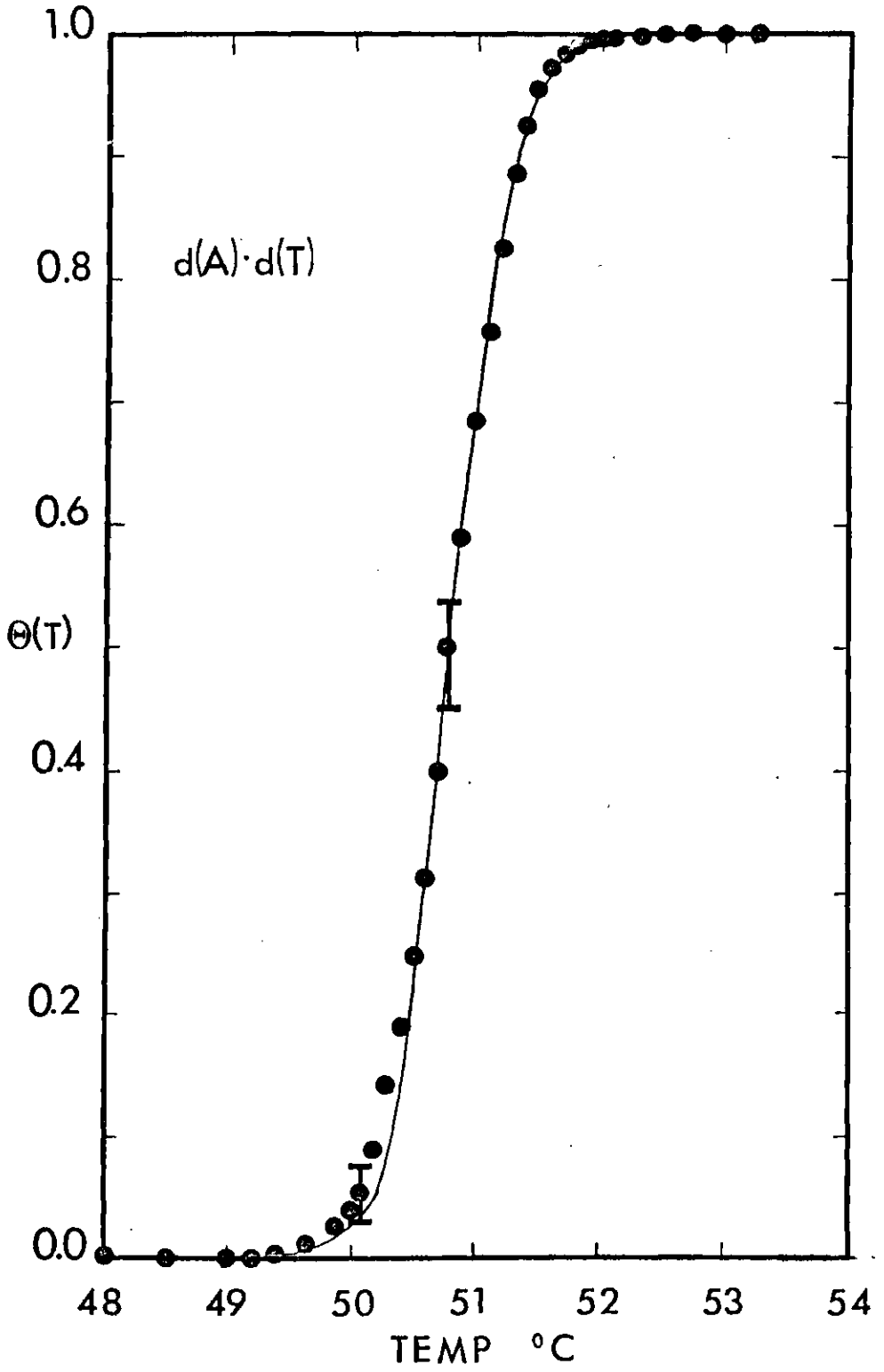


Figure 15. Helix-Coil Transition of d(A) · d(T).

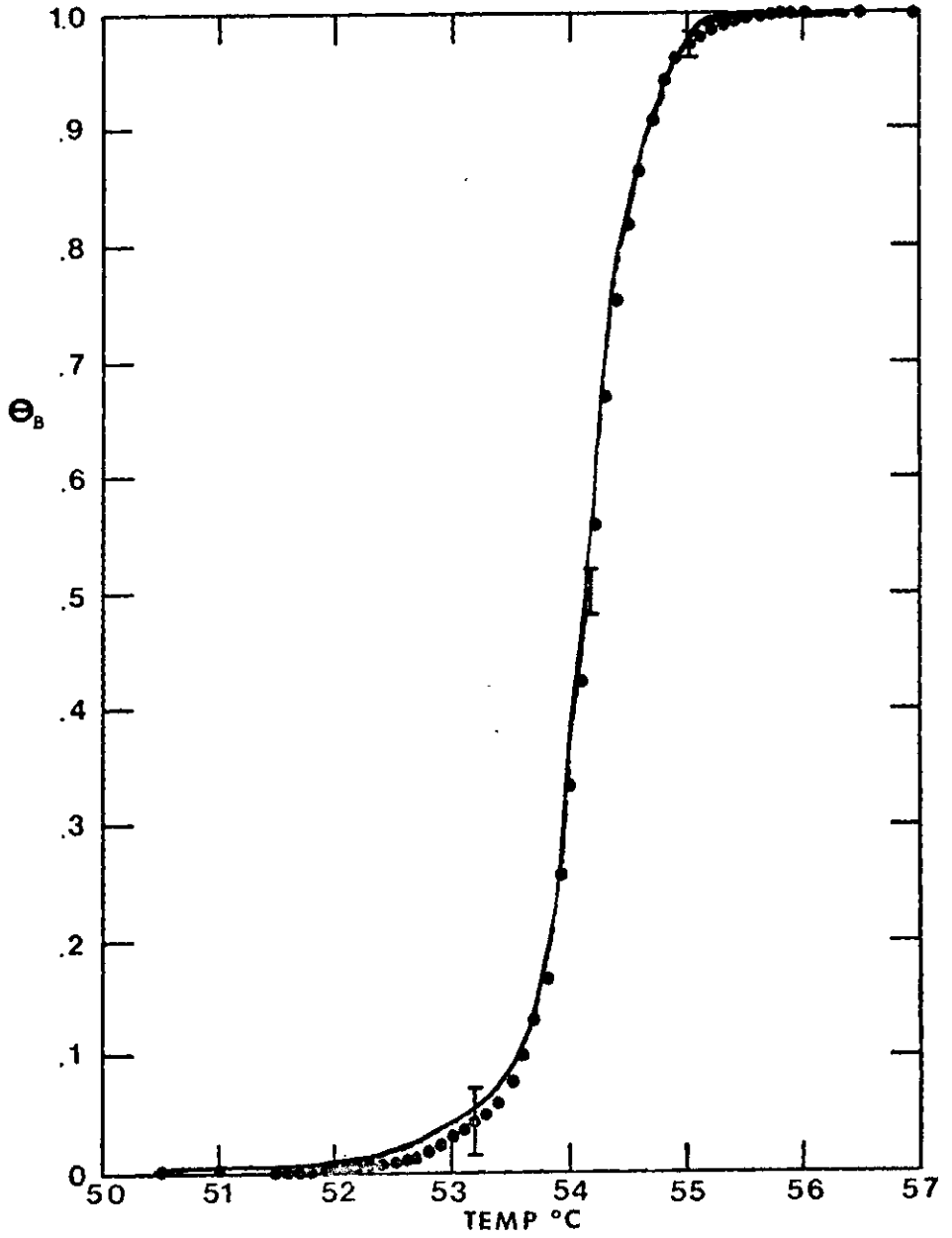


Figure 16. Helix-Coil Transition of d(AT) · d(AT).

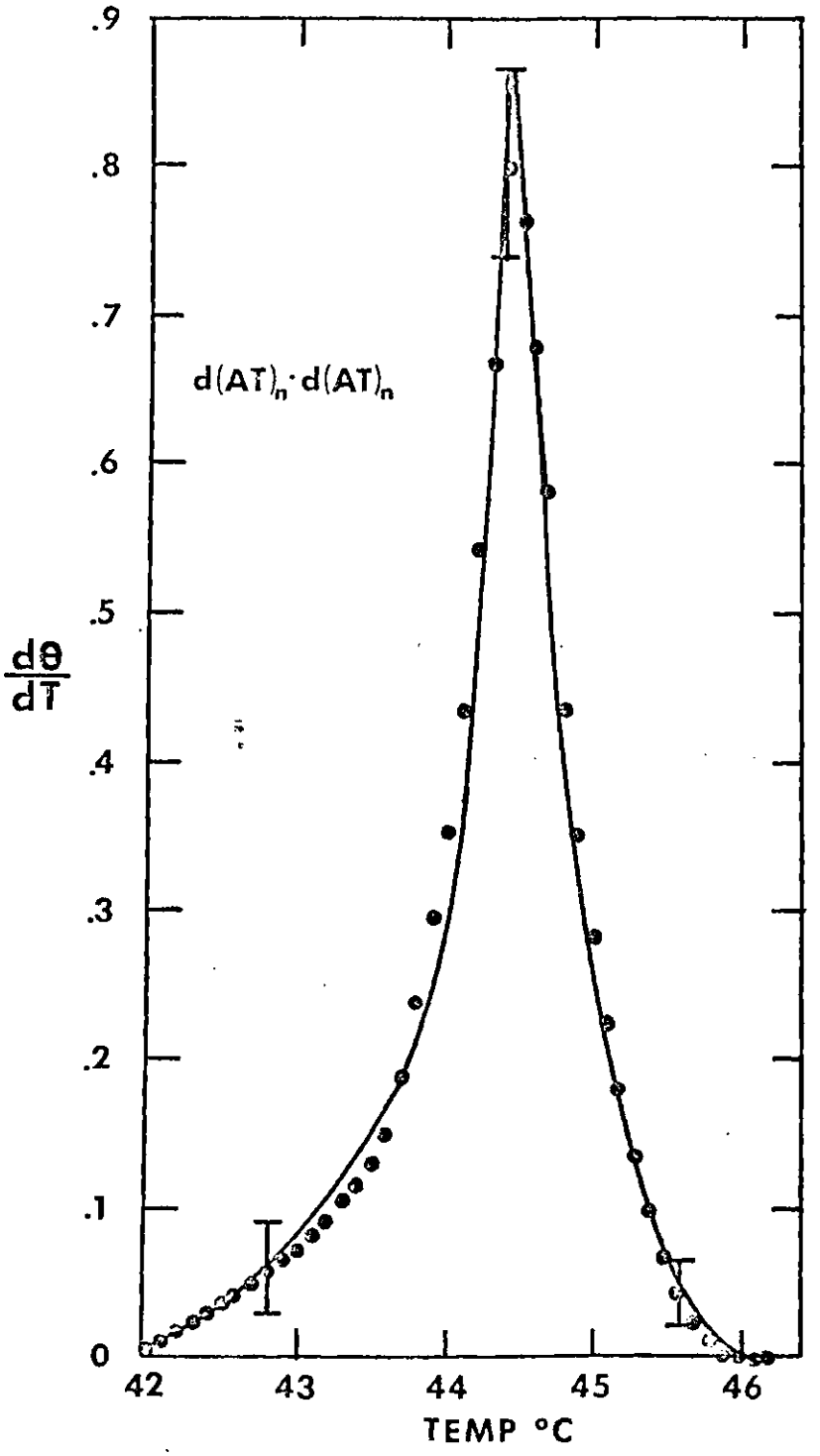


Figure 17. Derivative Curve of d(AT) · d(AT).

Table 2. Stability Parameters Evaluated From A-T DNA Melting Curves

Sodium Ion Molarity	K	$d(A)_n \cdot d(T)_n$ $G_s(11)$, kcal/mole	U_{11}	$d(AT)_n \cdot d(AT)_n$ $G_s(12$, kcal/mole	U_{12}
.003M	1.55	6.34	2.55	5.85	2.42
.015M	1.55	6.48	2.50	6.09	2.40
.045M	1.55	6.51	2.45	6.20	2.37

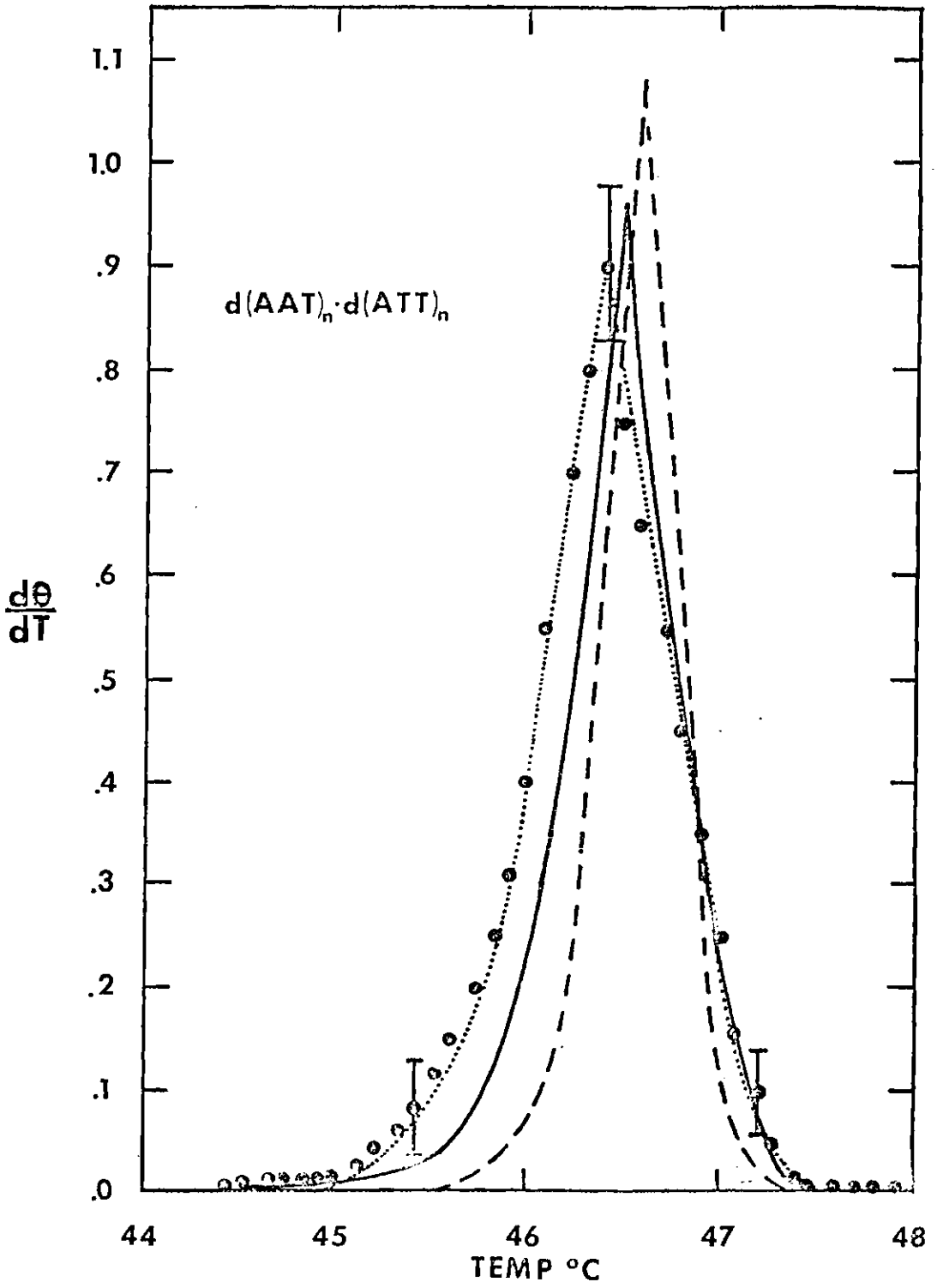


Figure 18. Experimental and Theoretical Melting Profiles of d(AAT) · d(ATT).

for d(AAT) · d(ATT) in 0.015 M sodium ion. The soild line represents the curve predicted by the parameters evaluated from d(A) · d(T) and d(AT) · d(AT); k=1.55, U_{12}=2.40, U_{11}=2.50. The dashed line is the curve predicted by parameters which fit the d(A) · d(T) and d(AT) · d(AT) curves when k was fixed at 1.70. Although the k=1.7 parameters were not the best ones, they produced curves within experimental error. It is observed that the solid line provides a reasonably good fit to the experimental curve. The dashed line predicts a curve substantially narrower than the one observed. This result again indicates that k=1.55 is the best value for the loop entropy constant. The agreement between theory and experiment for k=1.55 can be made even better by decreasing the U_{11} slightly. This will be discussed in detail in Chapter V.

GC DNAs

Poly d(G-C) · poly d(G-C) and poly d(G) · poly d(C) were melted using the same procedures as for the AT DNAs. The solvents used were .083 mM Na_2 EDTA + .417 mM Na H_2 PO_4 + .21 mM Na_2 H PO_4 and .83 mM Na_2 EDTA + 4.17 mM Na H_2 PO_4 + 2.1 mM Na_2 H PO_4 to give sodium ion concentrations of 10^{-3} M and 10^{-2} M respectively. Figure 19 shown the experimental data points for the transition of d(G-C) · d(G-C) in 10^{-3} M sodium ion. The solid line represents the best fit theoretical curve. It was generated with k=1.55 and

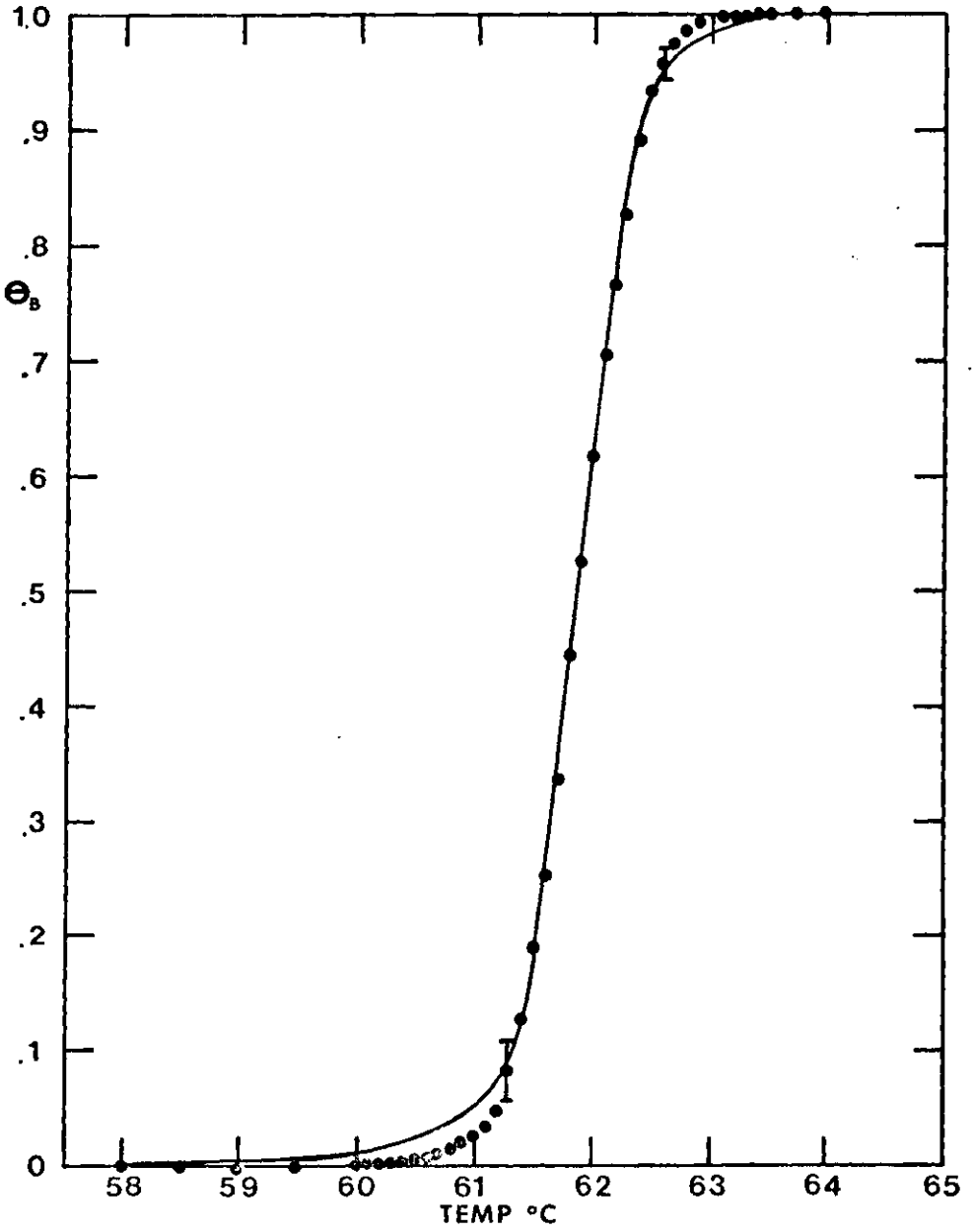


Figure 19. Helix-Coil Transition of d(GC) · d(GC).

U_{12}=2.50. Similar results were obtained for d(G) · d(C). Table 3 lists the evaluated parameters for both DNAs in both of the salt concentrations examined. Also listed in Table 3 are the stacking free energies evaluated from the relation between U_{12} and ΔG_s as shown in Equation (9). $\Delta G_s(12)$ is more negative than $\Delta G_s(11)$ in each of the solvents examined. This result is consistent with the observations that d(G-C) · d(G-C) has a greater Tm than d(G) · d(C) for a particular solvent.

Table 3. Stability Parameters Evaluated From G-C DNA Melting Curves

Sodium Ion Molarity	K	$d(G)_n \cdot d(C)_n$ $G_s(11)$, kcal/mole	U_{11}	$d(GC)_n \cdot d(GC)_n$ $G_s(12)$, kcal/mole	U_{12}
.01M	1.55	6.69	2.35	7.35	2.50
.001M	1.55	6.11	2.28	6.81	2.40

CHAPTER V

DISCUSSION

Introduction

One of the primary goals of a study of this nature is the understanding of the influence of base sequence on the conformational properties of DNA molecules. DNAs with known base pair sequences like synthetic DNAs are ideal candidates because of the relative ease in distinguishing the effects of different base pair sequences on various physical properties of the molecule.

Figures 15 and 19 illustrate the close agreement between the theory and experiment for the simpler AT and GC DNAs. Two observable differences between the theoretical and experimental transition curves are the actual and predicted melting temperature, Tm, and the width of the transition, i.e., the slope of the curve through the transition region. The experimental Tm is one of the parameters used in the theory so the predicted Tm is forced to be near the experimental Tm. However, as a result of the loop entropy factor being included in the calculations, the two Tm's are always different by .1 to .2 degrees centigrade with the predicted Tm being the higher of the two. Thus the amount of shift observed for different DNAs is an

indirect measure of the relative importance of looping. Table 4 lists the experimental and theoretical Tm's for all the DNAs examined in the various solvents used.

Figure 18 in Chapter 4 shows the close agreement between the experimental d(AAT) · d(ATT) and the transition predicted from the theory using the stacking parameters evaluated from d(A) · d(T) and d(A-T) · d(A-T) and the experimental Tm's of those same two DNAs. The difference is observed and predicted Tm's for d(AAT) · d(ATT) using those parameters is .2 degrees centigrade. Using the experimental d(A) · d(T) and d(AT) · d(AT) Tm's to predict the d(AAT) · d(ATT) Tm is incorrect however; instead those Tm's should be adjusted downward slightly to reflect the higher predicted Tm's for those two DNAs. When this is done, the predicted d(AAT) · d(ATT) Tm is only .1 degree centigrade higher than the observed experimental value.

Evaluation of Stacking Free Energies

Table 2 shows the stacking free energies for the AT DNAs and indicates that $\Delta G_S(11)$ is always more negative than $\Delta G_S(12)$ for a given salt concentration. This is consistent with the observed fact that d(A) · d(T) always melts at a higher temperature than d(A-T) · d(A-T) for a given salt concentration.

As mentioned in the preceding discussion, there are some discrepancies between the actual and predicted d(AAT) ·

Table 4. Actual and Predicted Tm's of the DNAs examined

	DNA	Actual Tm used in theory	Predicted Tm
.003 M Na+	d(A-T) · d(A-T)	29.20	29.30
	d(A) · d(T)	38.10	38.24
	d(AAT) · d(ATT)	32.55	32.75
.015 M Na+	d(A-T) · d(A-T)	44.40	44.50
	d(A) · d(T)	50.70	50.86
	d(AAT) · d(ATT)	46.40	46.60
.045 M Na+	d(A-T) · d(A-T)	54.15	54.20
	d(A) · d(T)	59.2	59.35
	d(AAT) · d(ATT)	55.55	55.85
10^{-2} Na	d(G-C) · d(G-C)	94.50	95.0
	d(G) · d(C)	82.7	82.85
10^{-3} Na	d(G-C) · d(G-C)	81.9	82.3
	d(G) · d(C)	61.85	62.0

d(ATT) melting profiles. This is not too surprising in light of several recent studies. Circular dichroism studies[9,45] on d(A) · d(T), d(AT) · d(AT), and d(AAT) · d(ATT) in solution indicated that the d(ApA) · d(TpT) dimer assumed different conformations in d(A) · d(T) and d(AAT) · d(ATT). Fiber diffraction studies[46,47] have shown that d(A) · d(T) exists in a B' conformation while d(AT) · d(AT) and d(AAT) · d(ATT) assume the B conformation under high humidity conditions. It seems reasonable to expect the stacking energy for d(ApA) · d(TpT) to change with conformation. Thus U_{11} evaluated from d(A) · d(T) is probably not appropriate for d(AAT) · d(ATT). We have observed that decreasing U_{11} by 0.08 in 0.003 M salt, 0.06 in .015 M salt and 0.03 M .045 M salt brings the theoretical curves into very good agreement with the d(AAT) · d(ATT) transitions as shown in Figure 17. This implies that stacking energy for the d(ApA) · d(TpT) dimer is less stabilizing when this dimer is in d(AAT) · d(ATT) than when in d(A) · d(T). Apparently less energy is lost by converting the d(ApA) · d(TpT) dimer from B' to the B form than by converting d(ApT) · d(ApT) from the B form to B'.

Table 3 shows the stacking energies for the GC DNAs. In this case the $\Delta G_s(12)$ is more negative than the $\Delta G_s(11)$, which again reflects the higher melting temperature for the alternating GC DNA, but is reversed for the situation for

the AT DNAs where $|\Delta G_s(11)| > |\Delta G_s(12)|$. Also, as shown in Table 4, the difference in actual and predicted Tm's is greater for the alternating GC than for the d(G) · d(C). These observations lead me to expect that when poly d(GGC) · poly d(GCC) is synthesized and characterized by melting studies that the $\Delta G_s(12)$ will be found to have a greater effect than $\Delta G_s(11)$ on the melting transition. This suggests that the alternating G-C base pair conformation changes to another conformation in the polymer d(GGC) · d(GCC). Further studies will be necessary to yield more information on the subject.

Evaluation of the Loop Entropy Exponent

The loop entropy term is included in the theory as a longer range interaction than the nearest neighbor interactions described by the U's and G's and is equal to the configurational entropy which remains in unbonded strands after connecting bonds are broken. This entropy depends on the size of the loops and the flexibility of the DNA strands and is usually incorporated into the theory with the use of Equation 14. Since the introduction of this idea by Zimm,[27] the evaluation of the loop entropy factor K has been the subject of several studies,[10,25,32] none of which has been conclusive. Calculations have indicated that K is between 1.5 and 2.0. A previous work[25] by one of the authors indicated $1.5 < K < 1.7$ for synthetic AT DNAs. Our

analysis gave K=1.55 ± .05 as that value of K which generated the theoretical curves which most closely matched the experimental curves for all three DNAs examined. If K was fixed at other values, the predicted Tm's were further from the actual Tm's, the slope through the transition for predicted curves got further from the actual slope, and the stacking free energies evaluated for the different DNAs did not reflect the higher or lower melting temperatures of those polymers.

The evaluated K=1.55 ± .05 value for the loop entropy factor is very close to the Jacobson-Stockmayer[48] value, 1.50, for an intersecting freely jointed polymer chain. This latter value is frequently employed as the standard value for the exponent of the loop entropy factor for matched loops. It seems reasonable to expect the physical factors not considered in the Jacobson-Stockmayer treatment to alter K for real DNA loops.

The loop entropy function e(m) is proportional to the probability that two DNA chains joined at one end will only meet at their free ends. For the periodic DNA polymers there are three factors which could alter the Jacobson-Stockmayer prediction of K=1.50: (1) the influence of the excluded volume of polymer segments on the probability that two chains joined at one end will only meet at their free ends, (2) the reduction of the configurational entropy of the strands due to restricted rotational angles

along the polymer bonds and solvent effects and (3) the reduction in the sliding degeneracy of loops due to a bias toward matched loops during melting. Theoretical calculations predict that factor (1) will increase K above the standard value.[49,50] The second factor will make internal loop formation easier and lower K. Delisi and Crothers[51] have calculated the loop entropy function for small loops with these two factors considered. When extrapolated to large loops their results are similar to the Jacobson-Stockmayer result. The third factor is important for the periodic DNAs. If the loop degeneracy was related to $(m)^{1/2}$ instead of m for d(A) · d(T) then the evaluated K will be higher than the actual K for matched loops. Similar arguments hold for the other two DNA polymers. One interpretation of my results is that the first two factors tend to cancel each other and the bias toward matched loop formation is small. Since quantitative estimates of these factors is uncertain, it is probably best to view $m^{-1.55}$ as the best empirical estimate for the loop entropy function of large loops.

The analysis employed in this work assumes that melting from the ends is negligible. The validity of this assumption is not certain. However, two results argue in its favor. I have not observed any difference between the melting curves of two d(A) · d(T) fractions estimated to be 4,000 and 6,000 base pairs long and theoretical studies

by Sture, Nordholm and Rice[53] indicate that melting from ends will be important for K 2 in the mismatching strand model of a homogeneous DNA. Our evaluated K is less than two.

Summary

The thermally induced helix-coil transitions of three AT DNAs d(A) · d(T), d(AT) · d(AT), d(ATT) · d(AAT) and two GC DNAs (d(G) · d(C), d(GC) · d(GC) were studied. The aim of this work was to evaluate thermodynamic parameters which govern DNA stability and to test the theoretical model employed in the analysis. The parameters evaluated from d(A) · d(T) and d(AT) · d(AT) should be consistent with those evaluated from d(AAT) · d(ATT). All DNA samples employed were high molecular weight fractions. Sedimentation velocity runs indicated they were > 4000 base pairs long. Experimental melting data was obtained by measuring the uv absorbence of the DNAs vs, temperature, T. This data was normalized to give the fraction of broken base pairs, θ, and $d\theta/dT$. The experimental melting curves were analyzed using the modified Ising model. In 0.003 M Na^+, a loop entropy exponent of k=1.55 provided an excellent fit to the transition curves of d(A) · d(T) and d(AT) · d(AT). The average stacking free energy for d(A) · d(T) was 6.34 kcal/mole. For d(AT) · d(AT) this value was 5.85 kcal/mole. The parameters evaluated from d(ATT) · d(AAT) were similar to, but not identical with those of the simpler DNAs.

APPENDIX

The figures found in this appendix represent the data taken and tabulated for the DNAs examined for this thesis that were not specifically talked about in the main body. The parameters evaluated for these DNAs can be found in Tables 1, 2, and 3.

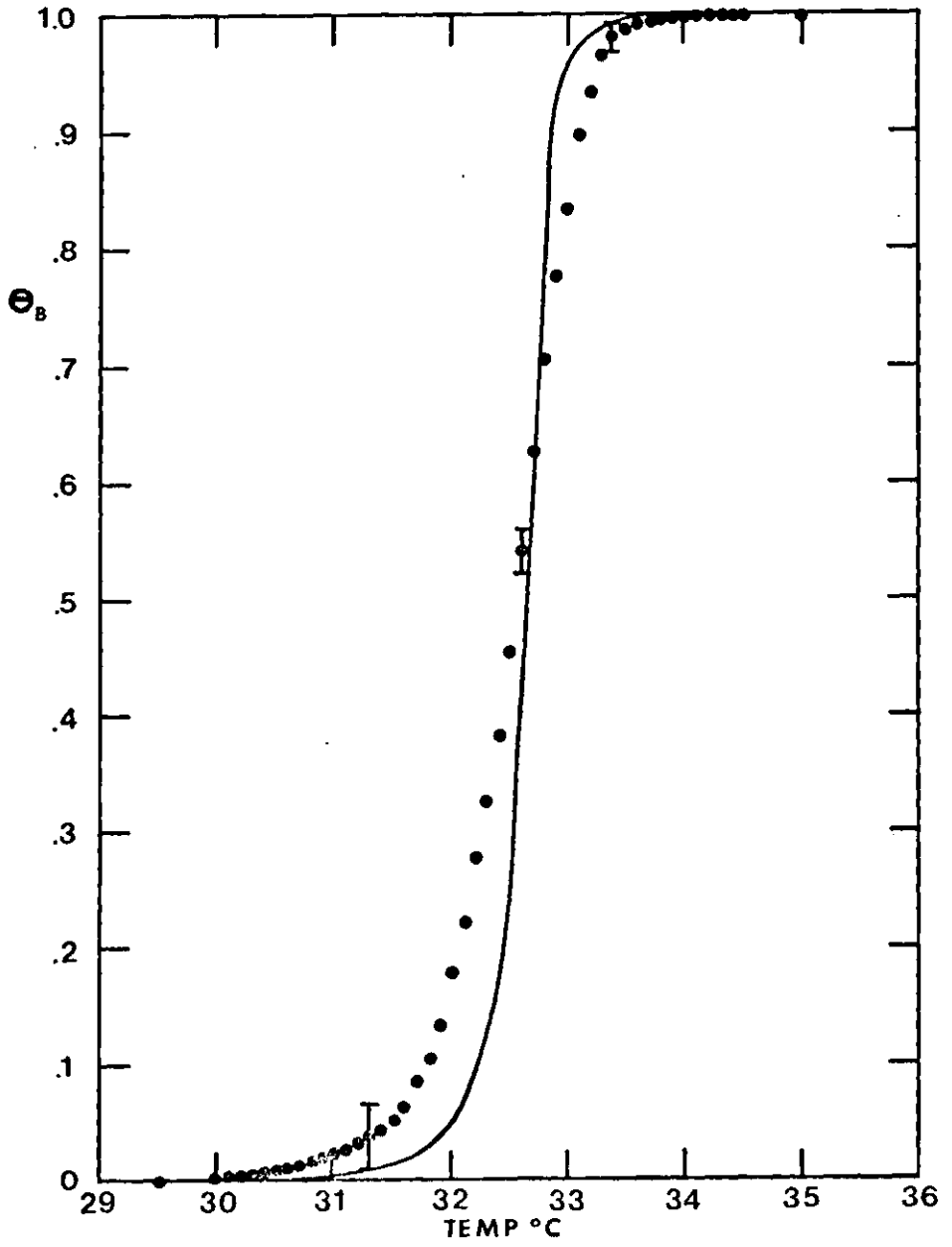


Figure 20. Helix-Coil Transition of d(AAT) · d(ATT) in .003 M Na+.

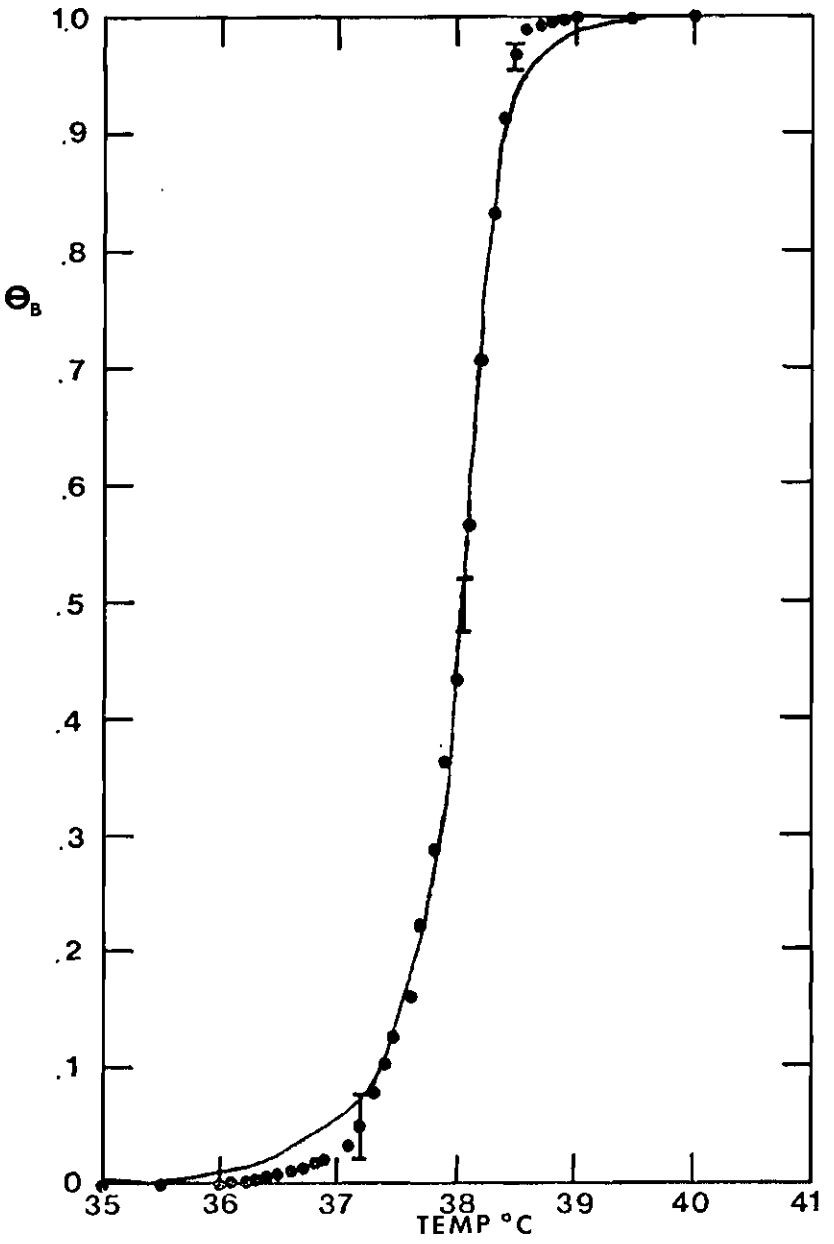


Figure 21. Helix-Coil Transition of d(A) · d(T) in .003 M Na+.

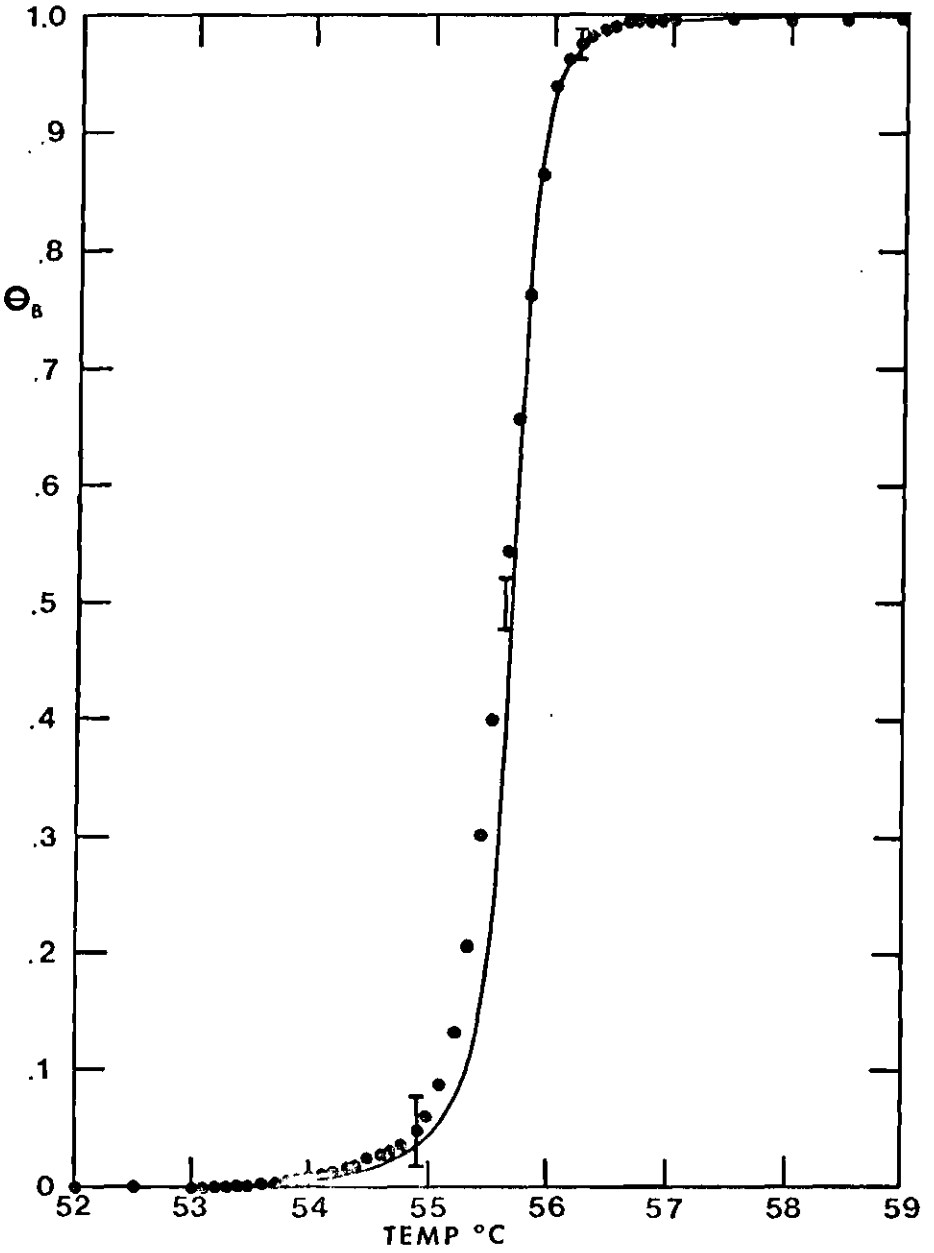


Figure 22. Helix-Coil Transition of d(AAT) · d(ATT) in .045 M Na+.

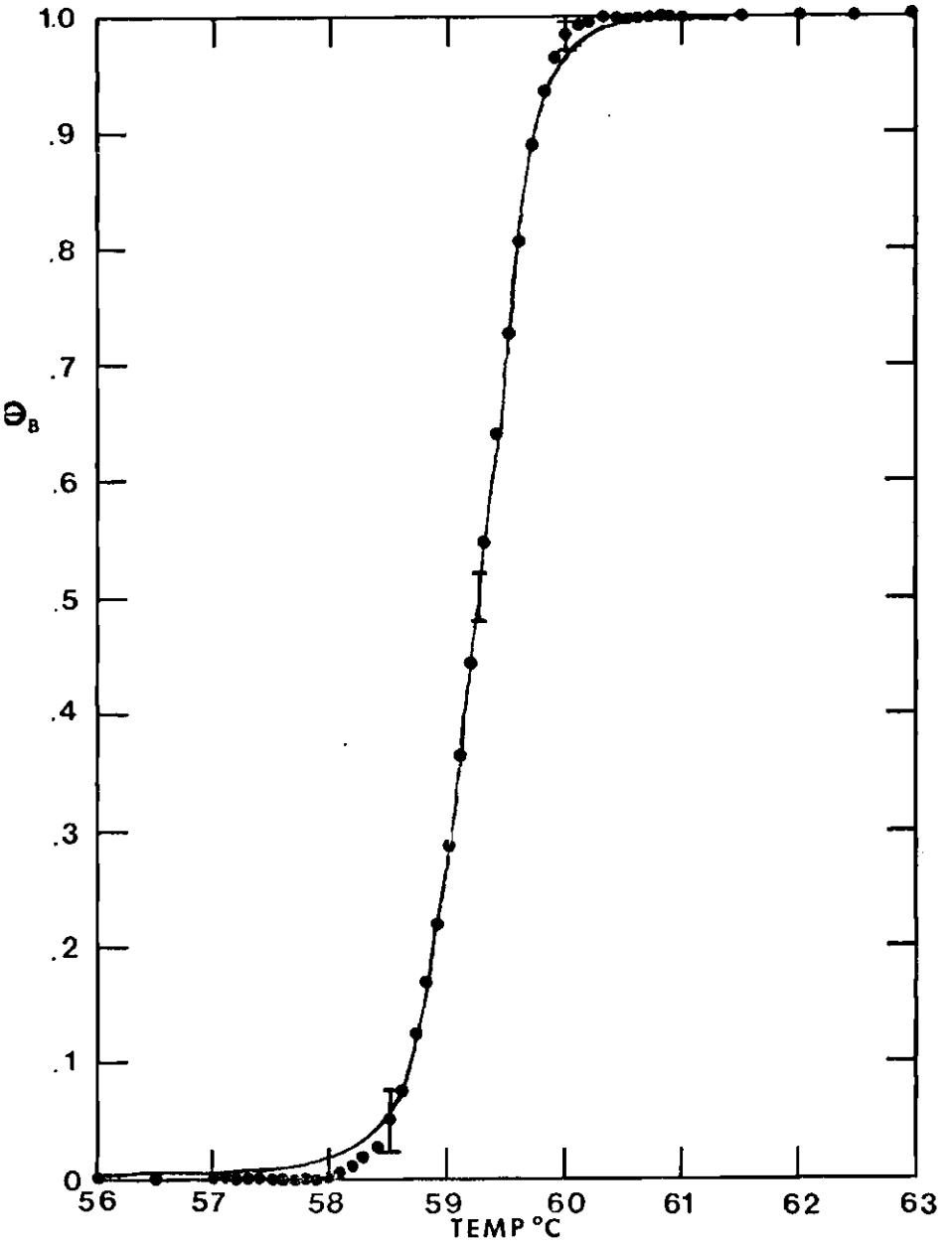


Figure 23. Helix-Coil Transition of d(A) · d(T) in .045 M Na+.

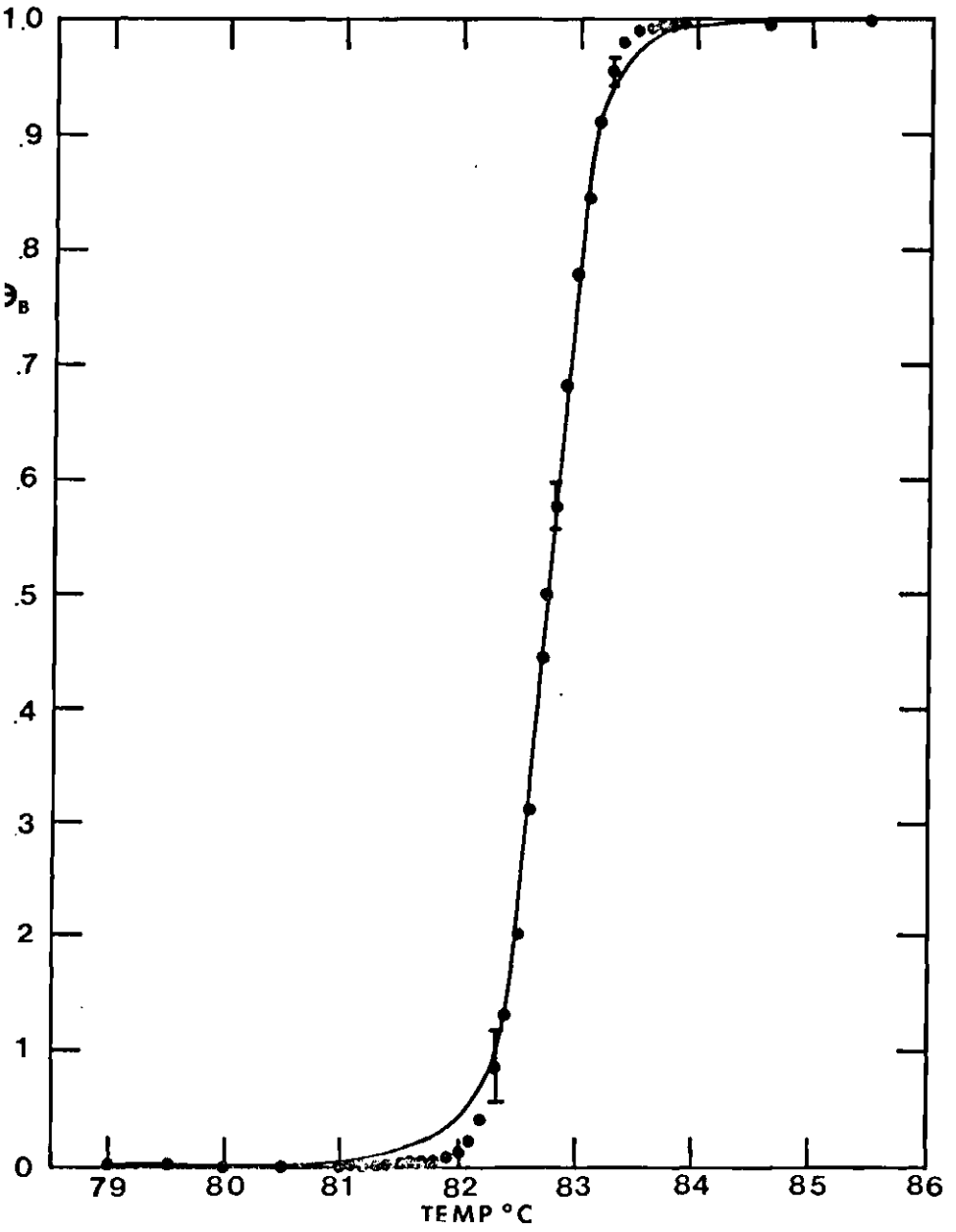


Figure 24. Helix-Coil Transition of d(G) · d(C) in .01 M Na+.

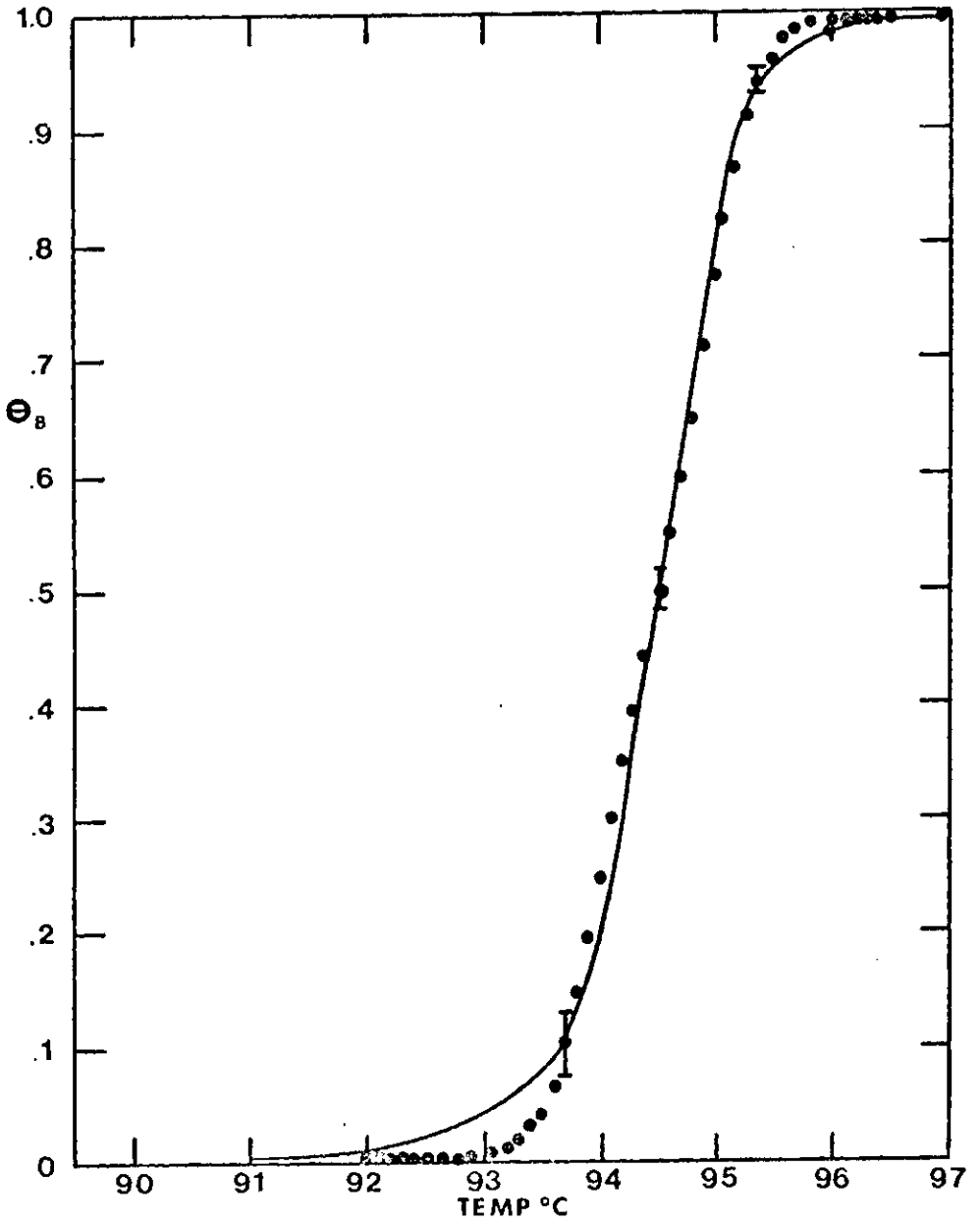


Figure 25. Helix-Coil Transition of d(GC) · d(GC) in .01 M Na+.

BIBLIOGRAPHY

1. J. D. Watson and F. H. C. Crick, Nature, 171, 737 (1953).

2. M. Messelson and F. W. Stahl, Proc. Natl. Acad. Sci., U. S., 44, 679 (1958).

3. A. A. Vedenov, A. M. Dykhne, and M. D. Frank-Kamenetska, Soviet Physics

4. I. E. Scheffler and J. M. Sturtevant, J. Mol. Biol., 42, 577 (1969).

5. R. D. Wells, E. Ohtsuka, and H. G. Khorna, J. Mol. Biol., 14, 222 (1956).

6. R. D. Wells and J. E. Blair, J. Mol. Biol., 27, 273 (1967).

7. F. M. Pohl and J. M. Jovin, J. Mol. Biol., 67,375 (1972).

8. D. M. Crothers, Acc. of Chem. Res., 2, 225, (1969).

9. D. M. Gray, R. L. Tatliff and D. L. Williams, Biopolymers, 12, 1233 (1973).

10. R. D. Blake, Biophysical Chemistry, 1, 24 (1973).

11. R. M. Wartell, Biopolymers, 11, 745 (1972).

12. I. V. Berestetskaya, M. D. Frank-Kamenetska and Y. S. Lazurkin, Biopolymers, 13, 193 (1974).

13. R. D. Wells, J. E. Larson, R. C. Grant, B. E. Shortle and C. R. Cantor, J. Mol. Biol., 54, 465 (1970).

14. R. B. Inman and R. C. Baldwin, J. Mol. Biol., 5, 172 (1962).

15. J. Marmur and P. Doty, J. Mol. Biol., 5, 109 (1962).

16. C. Schildkraut and S. Lifson, Biopolymers, 3, 195 (1965).

17. C. Schildkraut, J. Marmur, and P. Doty, J. Mol. Biol., 4, 430 (1962).

18. G. Felsenfeld and G. Sandeen, J. Mol. Biol., 5, 587 (1962).

19. G. Newell and E. W. Montroll, Rev. Mod. Phys., 25, 353 (1953).

20. D. M. Crothers and B. H. Zimm, J. Mol. Biol., 9, 1 (1964).

21. L. C. Klotz, Biopolymers, 7, 265 (1969).

22. E. W. Montroll and N. S. Goel, Biopolymers, 4, 855 (1966).

23. N. S. Goel and E. W. Montroll, Biopolymers, 6, 731 (1968).

24. R. M. Wartell and E. W. Montroll, Adv. Chem. Phys., 22 (1972).

25. R. M. Wartell, Biopolymers, 11, 745 (1972).

26. B. H. Zimm and J. K. Bragg, J. Chem. Phys., 31, 526 (1959).

27. B. H. Zimm, J. Chem. Phys., 33, 1349 (1960).

28. G. W. Lechman and J. P. McTague, J. Chem. Phys., 49, 3170 (1968).

29. A. A. Vedenov and A. M. Dykhne, Zh. Ekop. Teor. Fiz., 55, 357 (1968).

30. N. S. Goel, N. Fukuda, and R. Rein, J. Theoretical Biol., 18, 350 (1968).

31. D. M. Crothers, Biopolymers, 6, 1391 (1968).

32. D. Poland and H. A. Scheraga, Theory of the Helix Coil Transition in Biopolymers, Academic Press, New York, 1970.

33. J. F. Burd and R. D. Wells, J. Mol. Biol., 53, 435 (1970).

34. H. G. Khorna, H. Buchi, H. Ghosh, N. Gupta, T. M. Jacob, H. Kussel, R. Morgan, S. A. Narang, E. Ohtsuka, and R. D. Wells, Cold Spring Harb. Symp., 31, 39 (1966).

35. R. D. Wells, E. Ohtsuka and H. G. Khorna, J. Mol. Biol., 14, 221 (1965).

36. R. D. Wells, T. M. Jacob, S. A. Narang, and H. G. Khorna, J. Mol. Biol., 27,237 (1967).

37. R. D. Wells and R. M. Wartell, MTP International Review of Science, Biochemistry of Nucleic Acids, 6, 41 (1972).

38. D. A. Smith, A. M. Martinez and R. L. Ratliff, Anal. Biochem., 38, 85 (1970).

39. D. M. Gray, Biopolymers, 13, 2087 (1974).

40. J. Eigner and P. Doty, J. Mol. Biol., 12, 549 (1965).

41. R. L. Ratliff, D. L. Williams, F. N. Hayes. E. L. Martinez Jr., and D. A. Smith, Biochemistry 12, No. 24, 5005 (1973).

42. K. Olson, D. Luk, and C. L. Harvey, Biochem. Biophys. Acta., 277, 269 (1972).

43. R. D. Wells, J. Mol. Biol., 14, 221 (1965).

44. R. C. Grant, M. Kodama, R. D. Wells, Biochemistry 11, 805 (1972).

45. S. Arnott, Nucleic Acids Research, 2, 1493 (1975).

46. S. Arnott and E. Selsing, J. Mol. Biol., 88, 509 (1974).

47. E. Selsing, S. Arnott, and R. C. Ratliff, J. Mol. Biol., 98, 243 (1975).

48. G. H. Jacobson and W. H. Stockmayer, J. Chem. Phys., 18, 1600 (1950).

49. C. Domb, J. Gillis, and G. Wilmers, Proc. Phys. Soc. London, 85, 625 (1965).

50. M. Fisher, J. Chem. Phys., 45, 1469 (1966).

51. C. Delisi and D. M. Crothers, Proc. Natl. Acad. Sci., 68, 2682 (1971).

52. C. Delisi, Biopolymers, 12, 1713 (1973).

53. K. Sture, J. Nordholm and S. Rice, J. Chem. Phys., 59, 5605 (1973).

54. V. Bloomfield, D. Crothers, and I. Tinoco, Physical Chemistry of Nucleic Acids, Harper and Row, 1974.

VITA

Adrian L. Oliver was born in Mooselake, Minnesota on the fourteenth day of September in the year nineteen hundred and forty-six. Seventeen years later he graduated from Mankato High School, Mankato, Minnesota, in the top one third of a graduating class of 250 students. He then enrolled in and promptly flunked out of Mankato State College. The following four years were spent in the service of his country in the U. S. Army, during which time he became a commissioned officer and received several awards and commendations for excellence. Immediately upon separation from the service, Mr. Oliver enrolled in Augusta College, Augusta, Georgia, as a physics major. He took an average of 24 quarter hours per quarter and graduated with Honors, and the Physics Award and the Science Award eight quarters later. He then enrolled as a physics graduate student at the Georgia Institute of Technology where he took a Masters Degree in physics in 1972 and a Doctorate of Philosophy in 1976. He worked for eight months as a physics instructor for the Fernbank Science Center of the DeKalb County Board of Education and is presently employed with Varian Associates as a Marketing Engineer based in Atlanta, Georgia.